# Calculating Probabilities: Fundamentals of Bayesian Statistics

Ira Parsons, PhD

South Dakota State University
Department of Animal Science

# How do you know what you know, and how do you show what you know is really so?

Story time with Ira

# Introduction

- Three ways of drawing statistical inference
  - Frequentists
  - Likelihood
  - Bayesian
- Differences are sometimes controversial
- Modern scientists use the one that best fit their problem

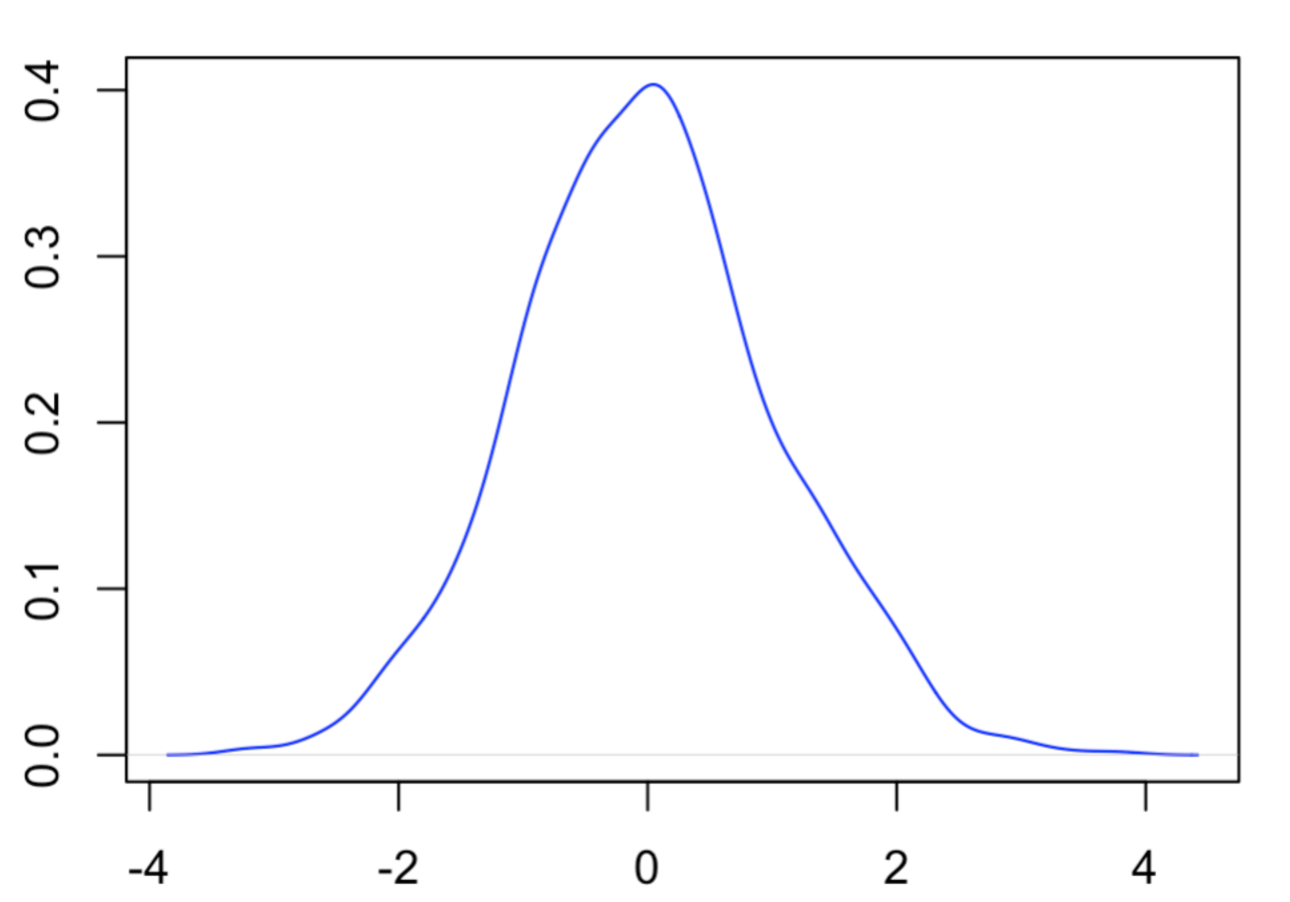
# Paradigms of Statistical Inference

- Three Paradigms of Inference
  - Frequentists
  - Likelihood
  - Bayesian
- Differences are sometimes controversial
- Scientists use the one that best fit their problem

# Bayes Theory

Thomas Bayes – 17$^{th}$ Century Mathematician and minister
Developed the theory of Inverse Probability

# Bayes Theory




Unobserved quantities are treated as Random variables

# Modeling sequence

## Design

Existing theory, scientific objectives, intuition

- Deterministic process model
- Collect observations

## Model specification

Diagram relationship → Write posterior distribution → Choose probability distribution

## Model implementation

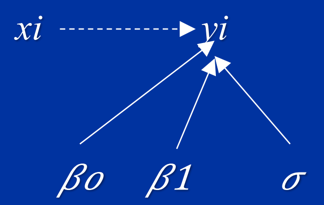
Write code → Implement on simulated data → Implement on real data

## Model evaluation and inference

Posterior predictive checks

- Probabilistic inference
- Model selection/ Averaging

# Models




$$g(\beta, xi) = \beta 0 + \beta 1\ xi$$

Priors

Beta0 = dnorm(0, 0.001)
Beta1 = dnorm(0, 0.001)
Tau = dgamma(0.001, 0.001)
Sigma_sq = 1/tau

```
model {

  # priors
  beta0 ~ dnorm(0,.001)
  beta1 ~ dnorm(0,.001)
  tau ~ dgamma(.001, .001)
  sigma_sq  <- 1/tau

  # likelihood
  for(i in 1:n) {
    mu[i] <- beta0 + beta1*x[i]
    y[i] ~ dnorm(mu[i], tau)
  }
}
```

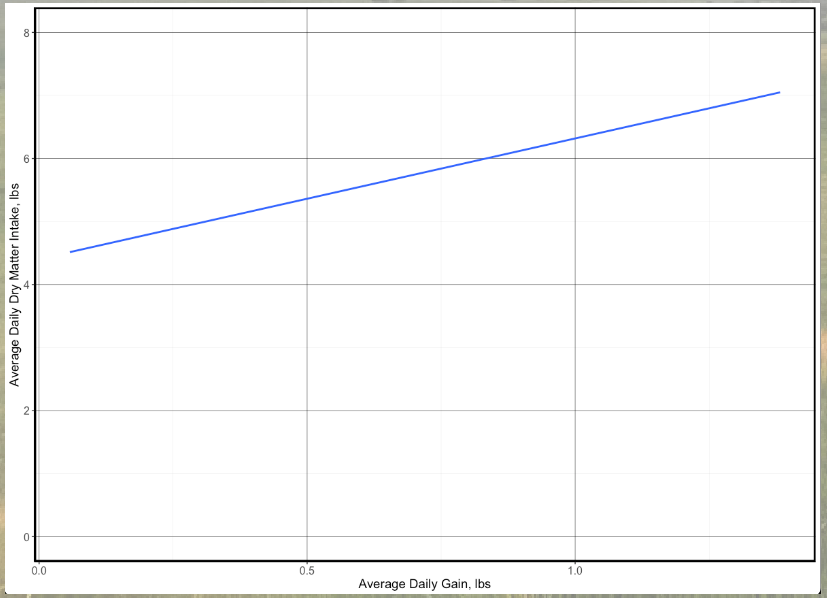
Bayesian Credible Intervals

# Deterministic process model

**Process model**

$g(\beta, xi) = \beta_0 + \beta_1 xi$

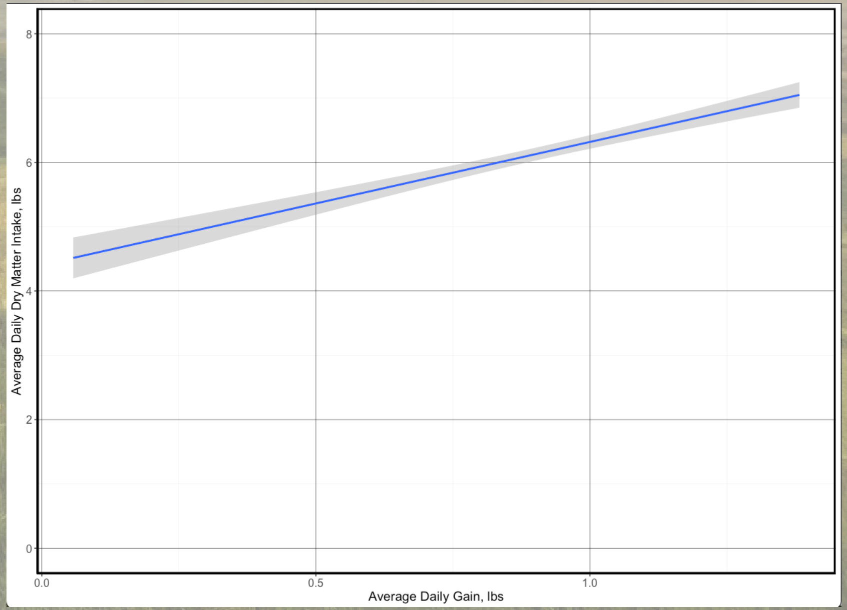
DMI ~ 4.5 + 1.91x




**Variance model**

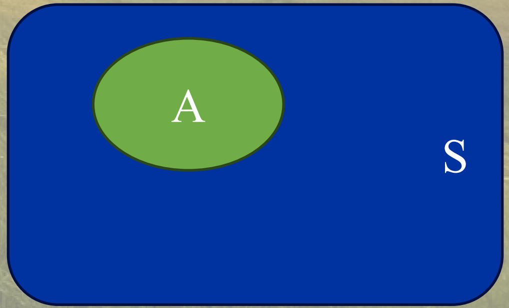
$g(\beta, xi) = \beta_0 + \beta_1 xi + \varepsilon$

DMI ~ 4.5 + 1.91x + variance



# Conditional Probability

Bayesian models



# Random variables

World is divided into things that are observed and things that are unobserved

1. Bayesian treat all unobserved quantities as *random variables*
2. Values of *random variables* are governed by chance
3. Probability distributions quantify "governed by chance"
   *Where chance occurs*

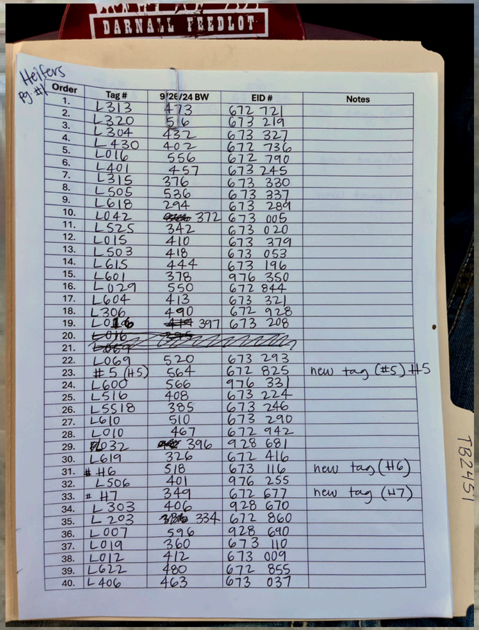

South Dakota
State University
Department of Animal Science

# Random variables

- Model parameters
- Latent states
- Missing data
- Predictions and forecasts
- Observations

# Random variables

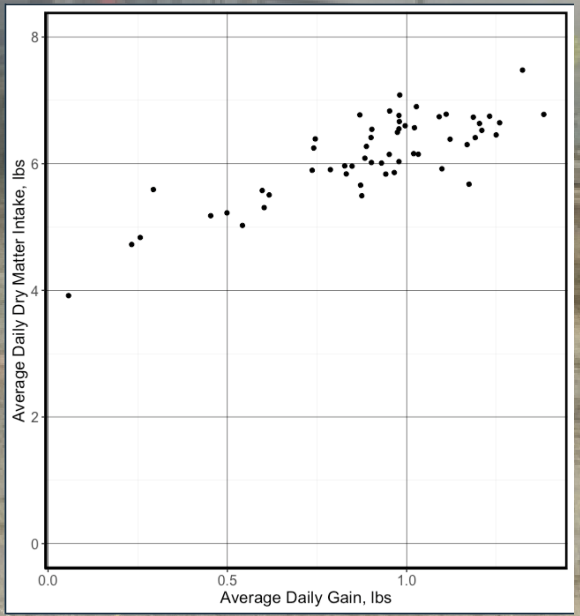
## Collected data



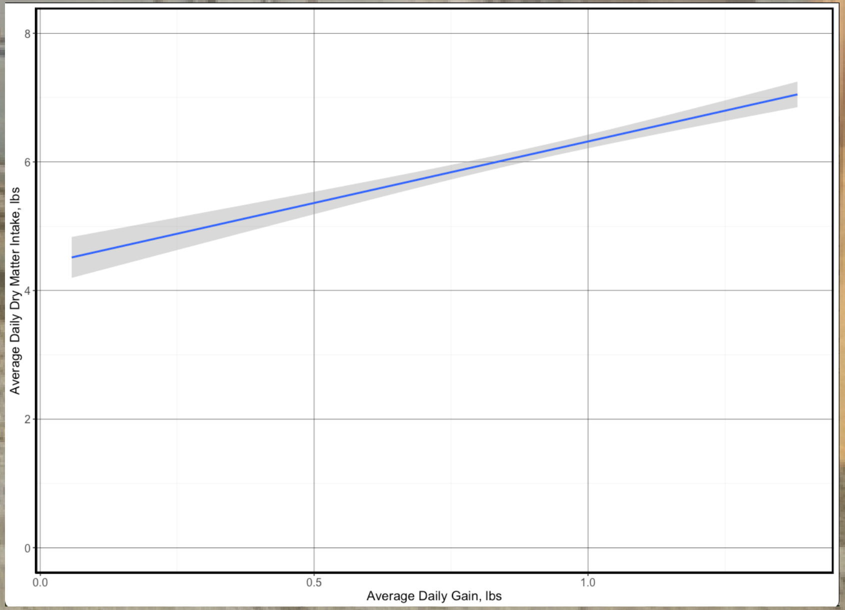

Heifers
pg. #1

| Order | Tag # | 9/26/24 BW | EID # | Notes |
|---|---|---|---|---|
| 1. | L313 | 473 | 672 721 | |
| 2. | L320 | 516 | 673 219 | |
| 3. | L304 | 432 | 673 327 | |
| 4. | L430 | 402 | 672 736 | |
| 5. | L016 | 556 | 672 790 | |
| 6. | L401 | 457 | 673 245 | |
| 7. | L315 | 376 | 673 330 | |
| 8. | L505 | 536 | 673 337 | |
| 9. | L618 | 294 | 673 289 | |
| 10. | L042 | ~~[illegible]~~ 372 | 673 005 | |
| 11. | L525 | 342 | 673 020 | |
| 12. | L015 | 410 | 673 379 | |
| 13. | L503 | 418 | 673 053 | |
| 14. | L615 | 444 | 673 196 | |
| 15. | L601 | 378 | 976 350 | |
| 16. | L029 | 550 | 672 844 | |
| 17. | L604 | 413 | 673 321 | |
| 18. | L306 | 490 | 672 928 | |
| 19. | L016 | ~~414~~ 397 | 673 208 | |
| 20. | ~~L016~~ | ~~395~~ | | |
| 21. | ~~L069~~ | | | |
| 22. | L069 | 520 | 673 293 | |
| 23. | #5 (H5) | 564 | 672 825 | new tag (#5) H5 |
| 24. | L600 | 566 | 976 331 | |
| 25. | L516 | 408 | 673 224 | |
| 26. | L5518 | 385 | 673 246 | |
| 27. | L610 | 510 | 673 290 | |
| 28. | L010 | 467 | 672 942 | |
| 29. | L032 | ~~[illegible]~~ 396 | 928 681 | |
| 30. | L619 | 326 | 672 416 | |
| 31. | # H6 | 518 | 673 116 | new tag (H6) |
| 32. | L506 | 401 | 976 255 | |
| 33. | # H7 | 349 | 672 677 | new tag (H7) |
| 34. | L303 | 406 | 928 670 | |
| 35. | L203 | ~~[illegible]~~ 334 | 672 860 | |
| 36. | L007 | 596 | 928 690 | |
| 37. | L019 | 360 | 673 110 | |
| 38. | L012 | 412 | 673 009 | |
| 39. | L622 | 480 | 672 855 | |
| 40. | L406 | 463 | 673 037 | |

## Fixed variables




## Random βeta's



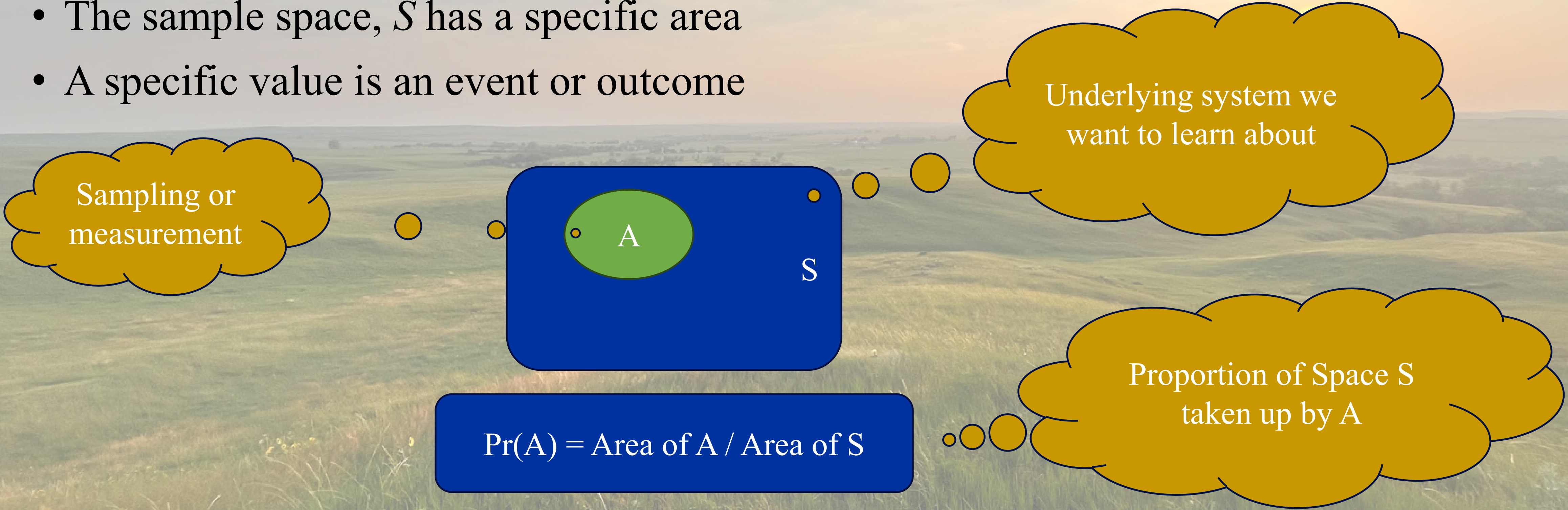
# Three rules of probability

1. Conditional probability
2. Law of total probability
3. Chain rule of probability

Think proportions of groups, subgroups, and contingencies.

# Defining the sample space

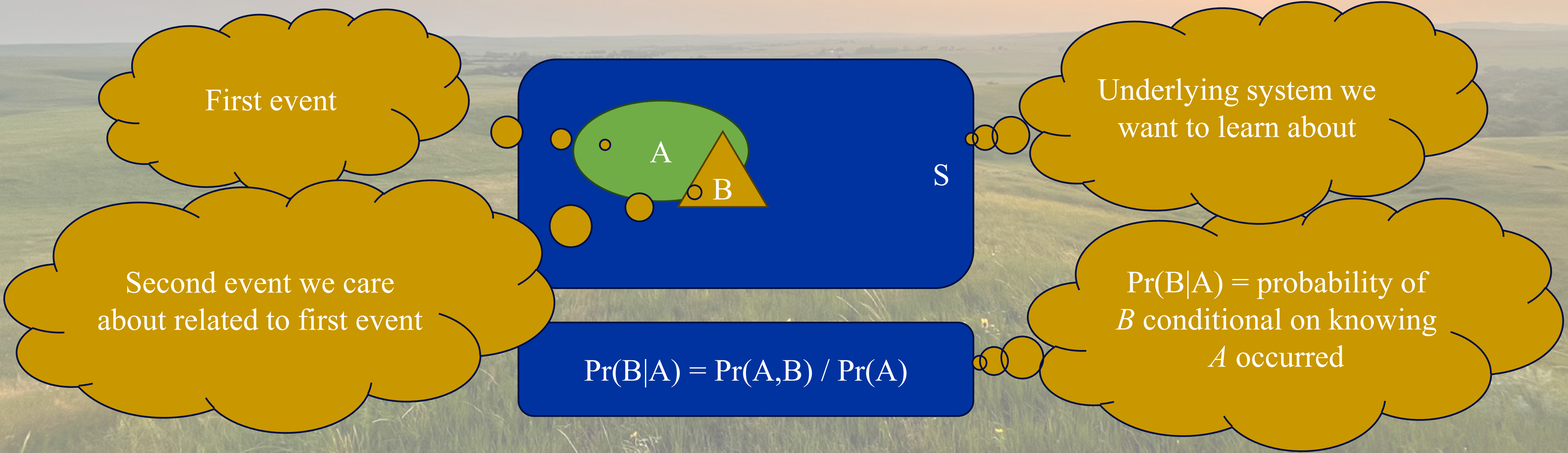
- The set of all possible values of a random variable
- The sample space, *S* has a specific area
- A specific value is an event or outcome



# Conditional Probability

*Conditional probability:* the probability of an event given that we know another event has occurred.

- What is the probability of event *B* , Given we know Event *A* has occurred

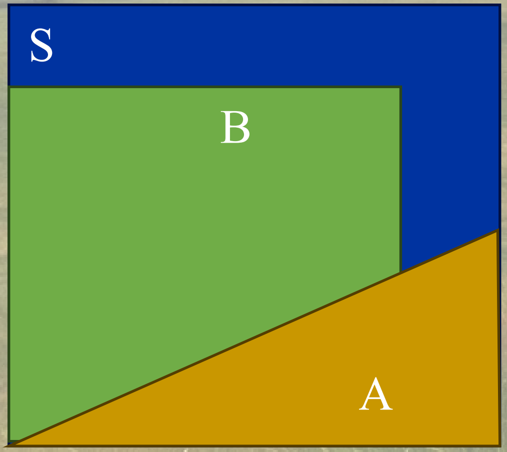


# Independence

Event A and B are *independent* if the occurrence of A does not tell us anything about B

$\Pr(A|B) = \Pr(A)$

$\Pr(B|A) = \Pr(B)$




Pr(A|B) = area of A and B / area of B

Pr(A|B) = area of A / area of S

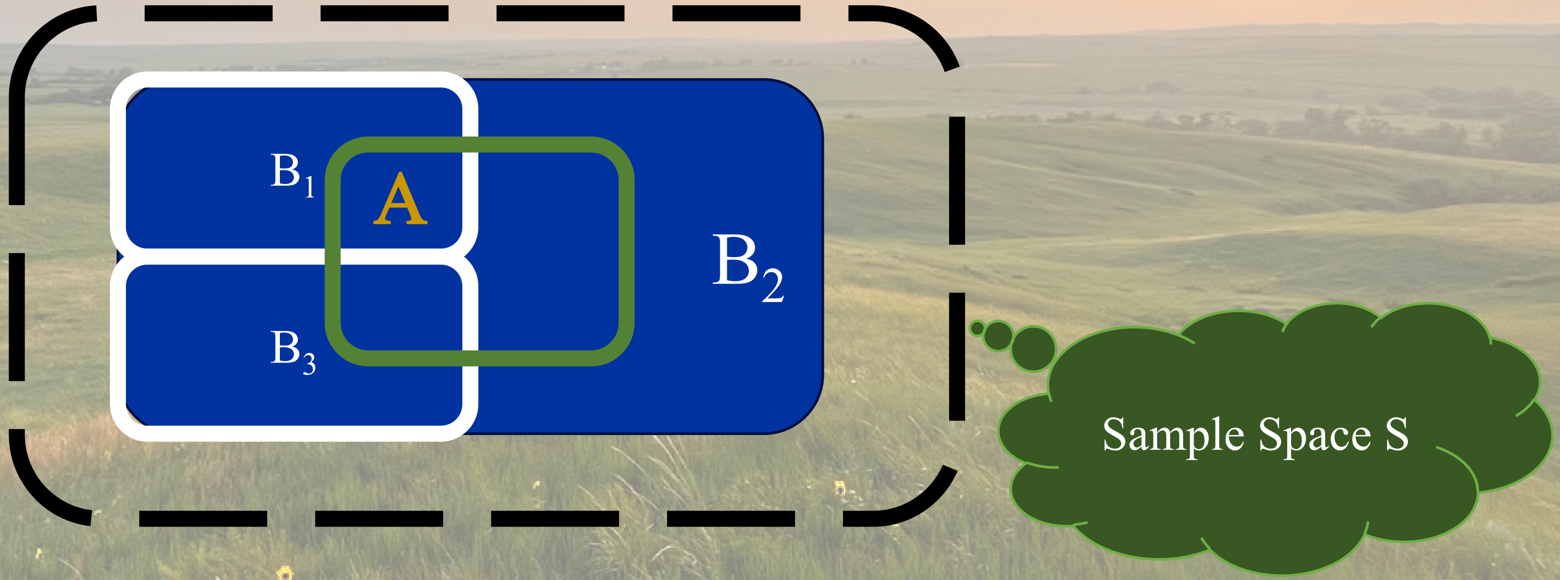
Can be rearranged

$\Pr(A,B) = \Pr(A|B)\Pr(B)$
$\Pr(A,B) = \Pr(B|A)\Pr(A)$

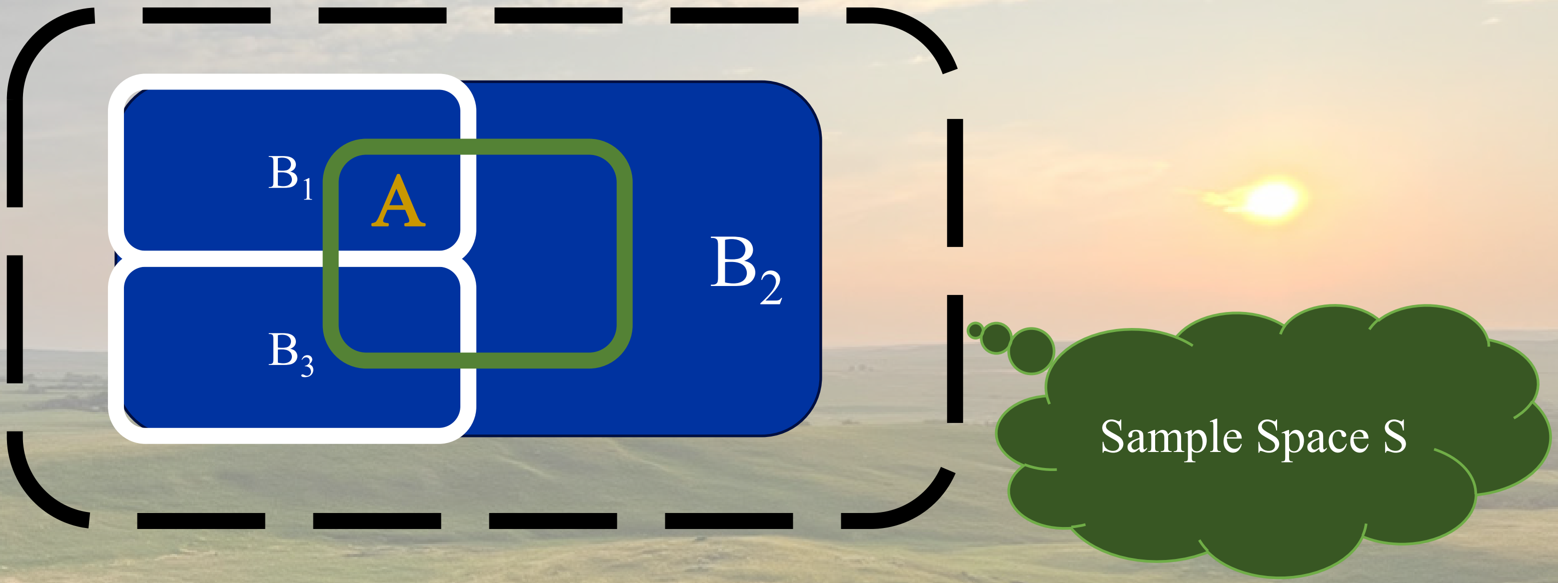
# Law of Total Probability

Pr(*A*) is unkown, but can be calculated using the known probabilities of several events

$B_n$: n = 1,2,3,… define the entire sample space S



# Law of Total Probability




Rearranging the expression of conditional probability

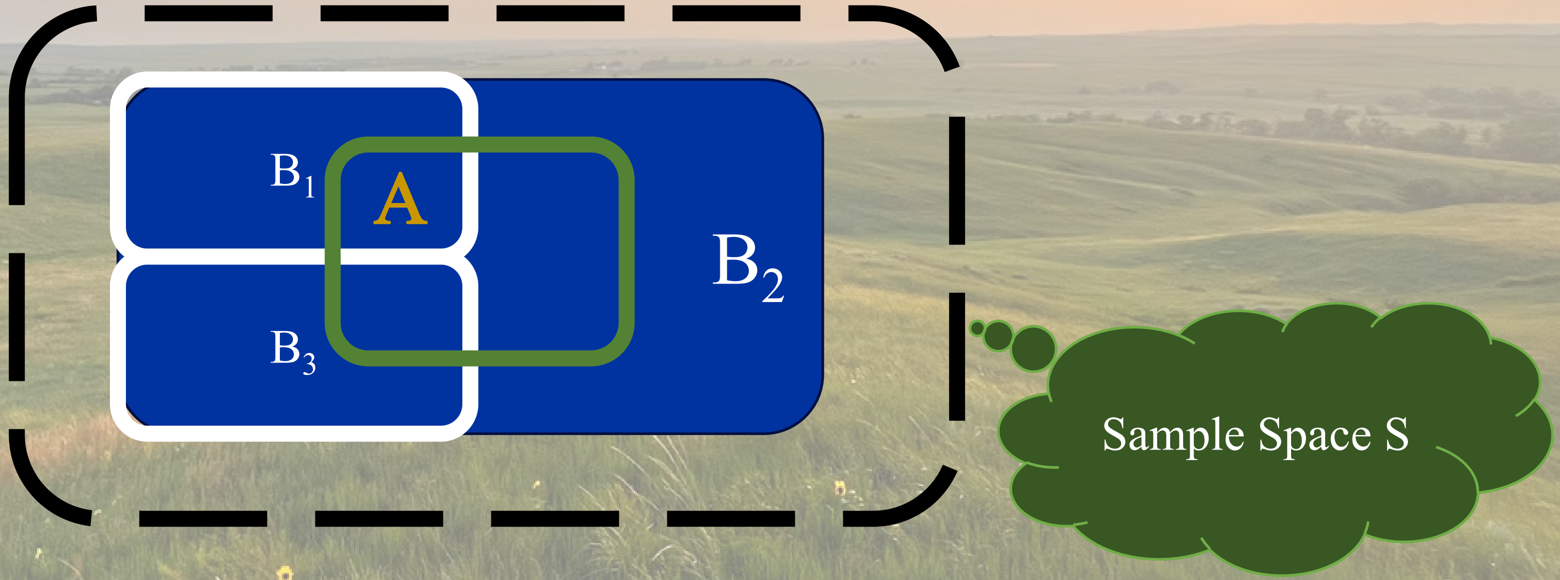
$Pr(A,B) = Pr(A|B)Pr(B)$

$Pr(A,B) = Pr(B|A)Pr(A)$

# Probability of Event A?

$\Pr(A) = \sum_n \Pr(A|B_n)\Pr(B_n) = \sum_n \Pr(A,B_n)$ discrete case

$\Pr(A) = \int \Pr(A|B)\Pr(B)\ B = \int \Pr(A,B) B$ continuous



# The Chain Rule of Probability

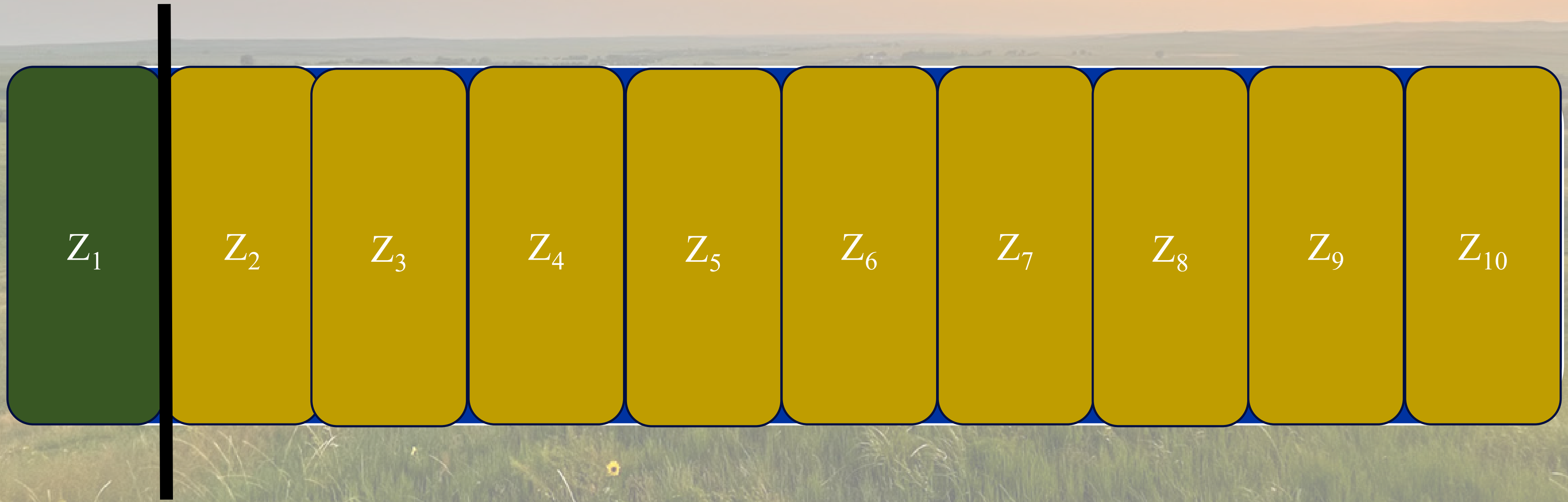
The chain rule of probability allows writing joint distributions as a product of conditional distributions.

$$Pr(z1, z2\ldots.,zn) = Pr(zn|zn-1,\ldots z1)Pr(zn-1|zn-2..,z1)Pr(z2|z1)Pr(z1)$$

- Z’s can be scalars or vectors
- Sequence does not matter
- Choose a sequence that makes sense

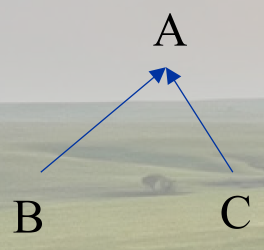
# Chain Rule of Probability

Pr(z1, z2….,zn) = Pr(zn|zn-1,…z1)Pr(zn-1|zn-2..,z1)Pr(z2|z1)Pr(z1)



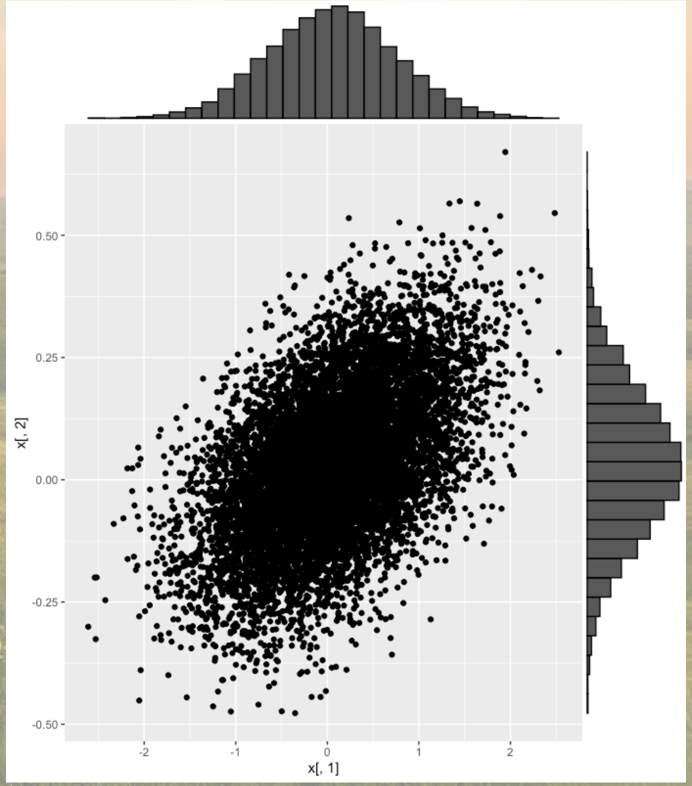
# Factoring joint probabilities

Directed Acyclic Graph



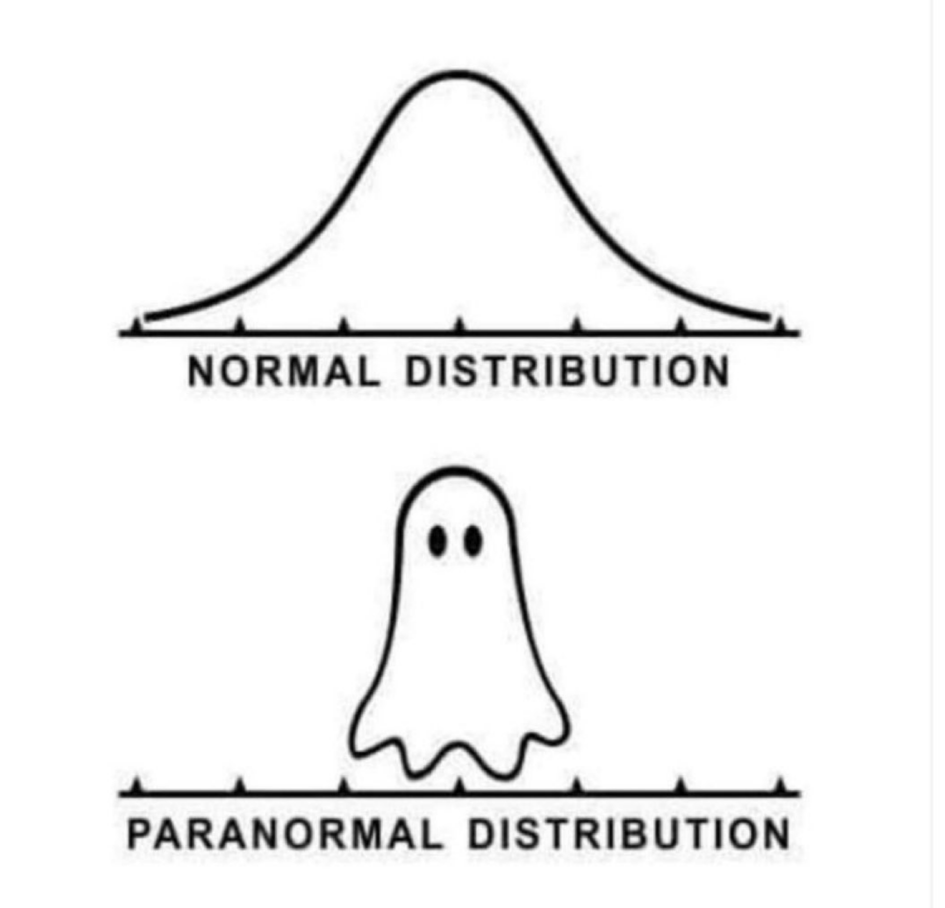

Represents [A|B,C][B][C]

- DAGs (Bayesian Networks) specify how joint distributions are factored into conditional distributions
- Nodes at the heads of arrows must be on the left side of conditioning symbols
- Nodes at tails must be on the right side of conditioning symbols
- Any node without an arrow leading to it must be expressed unconditionally



# Probability Distributions

A probability implies a distribution



# What we need to know

- Probability distribution are our toolbox for fitting models to data and representing uncertainty
- Moments are how we summarize probability distributions
- Every distribution is supported by underlying data
- The data **type** defines the support for the distribution

# Consider a Linear Function

$Y = mx + b$

- $y = f(x)$ is a function of x, with fixed values $m$ and $b$ each value of $x$ gets mapped to as single $f(x)$
- $x$ is our variable of interest

**Random Variables**

- **Sample space** encompasses all possible outcomes from a **random process**
- A **random variable** is a function from a particular sample space to real numbers

# Probability distribution components

| Probability model | Random variable support | Parameters | Moments |
|---|---|---|---|
| Normal | Real numbers | $u, \sigma^2$ | $u, \sigma^2$ |
| Lognormal | Positive real numbers | $\alpha$ mean of log of $z$<br>$\beta$ the standard deviation of the log of $z$ | $u, \sigma^2$ |
| Gamma | Positive real numbers | $\alpha$ = shape, $\beta$ = rate | $u, \sigma^2$ |
| Beta | Real numbers [0,1] or (0,1) | $\alpha$, $\beta$ | |
| Bernoulli | 0 or 1 | $\phi$ probability that random variable equals 1<br>$\phi = u$ | $u = \phi$<br>$\sigma^2 = \phi(1-\phi)$ |
| Binomial | Counts in two categories with upper bound | $n$ number of trials<br>$\phi$ probability of success | $u = n\phi$<br>$\sigma^2 = n\phi(1-\phi)$ |
| Negative binomial | Counts | $\lambda$ the mean number of occurences<br>$k$ dispersion parameter | $u = \lambda$<br>$\Sigma^2 = \lambda + \lambda^2 / k$ |

South Dakota State University
Department of Animal Science

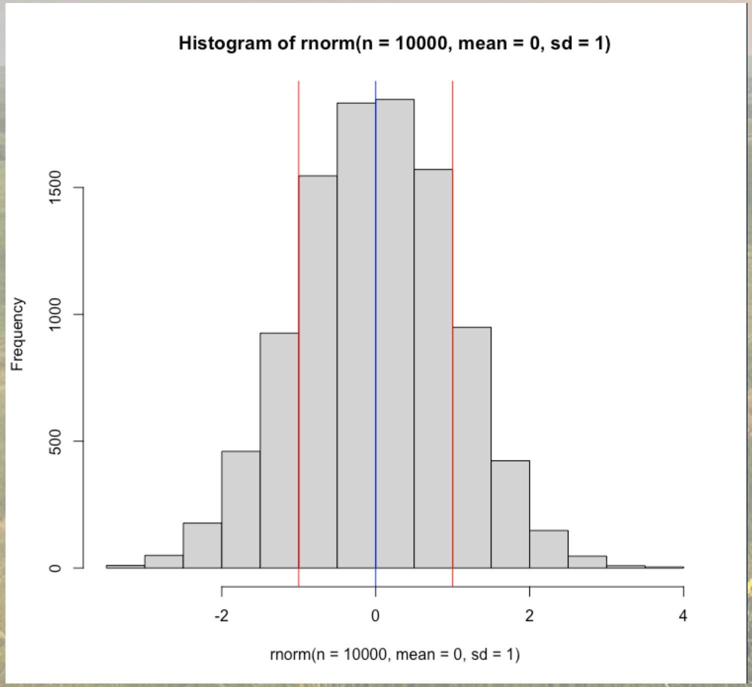
# Example Processes

| Item | Pregnancy check cows | Weaning heifers |
|---|---|---|
| **Random process** | Pregnancy check cows | Body mass |
| **Possible outcomes** | Pregnant or Not-Pregnant | Any amount of mass |
| **Random variable** | $X$ = number of pregnant cows | $Y$ = amount of body mass |
| **Support** | Sx = {0,1} | $Sy$: $y > 0$ |
| **Possible Probabilities of Interest** | Pr(Pregnant) = Pr(X=1) | Pr(>500lb heifer) = Pr(Y > 500) |

# Frequency distributions

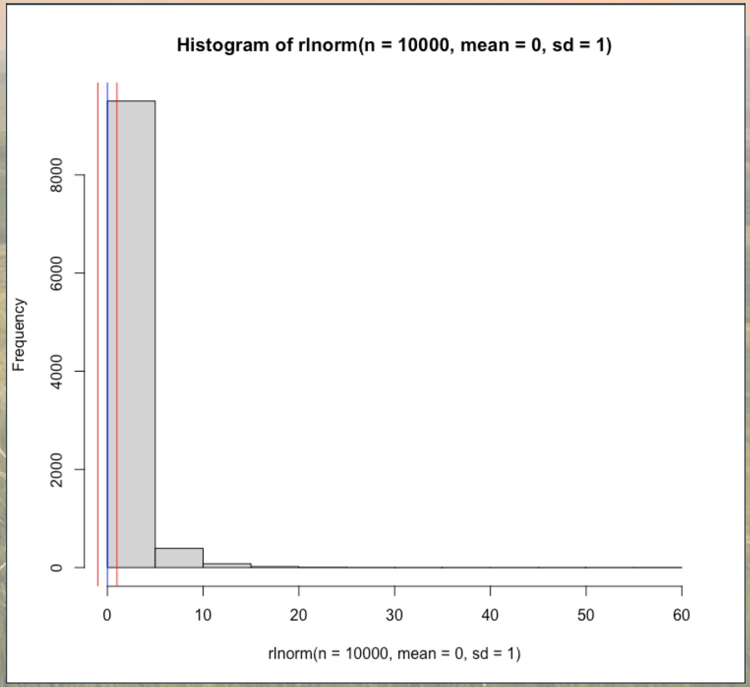
## Normal Distribution

```
hist(rnorm(n = 10000, mean = 0, sd = 1))
```




## Log-Normal Distribution

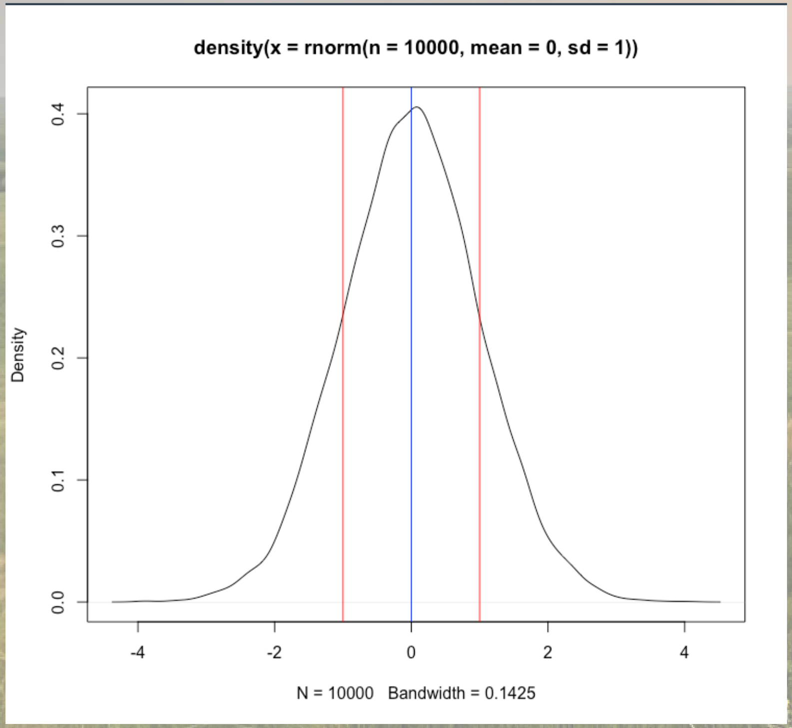
```
hist(rlnorm(n = 10000, mean = 0, sd = 1))
```



# Probability distributions

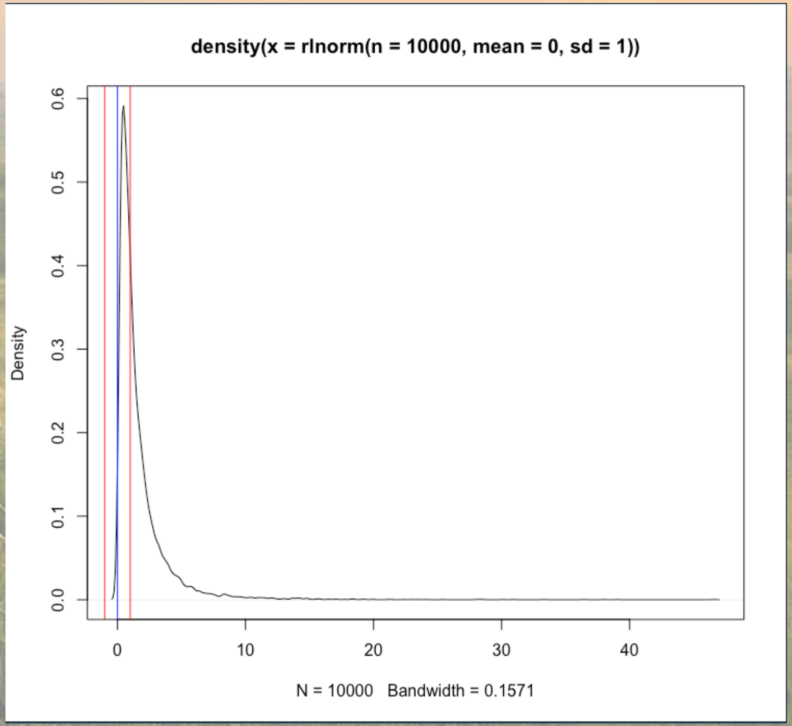
## Normal distribution

`plot(density(rnorm(n=10000, mean=0, sd=1)))`



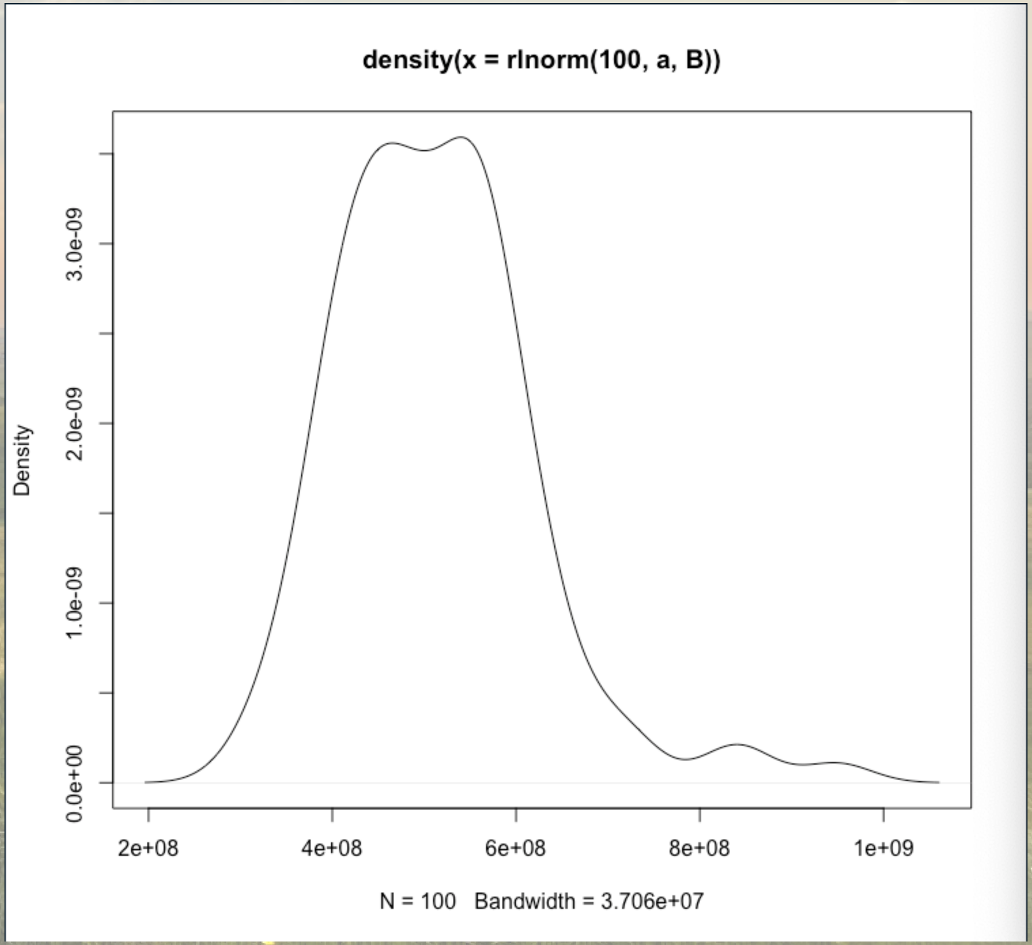

## Log-Normal distribution

`plot(density(rlnorm(n=10000, mean=0, sd=1)))`



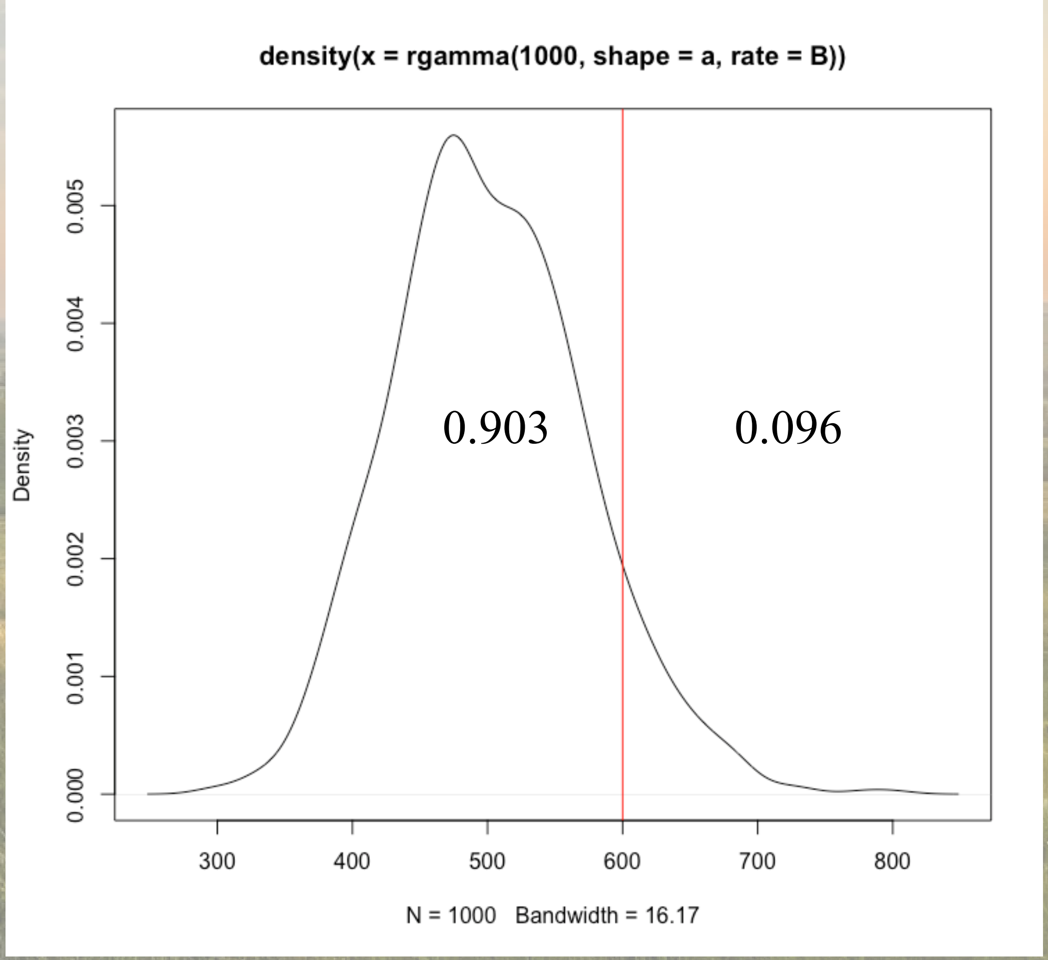
# Steer body-weight

```
u=500
sd=50
a = log(u) - 0.5*log((sd^2 + u^2)/u^2)
B = sqrt(log((sd^2 + u^2)/u^2))
plot(density(rlnorm(10000,a,B)))
```



# Probability of observeing a steer > 600 lbs

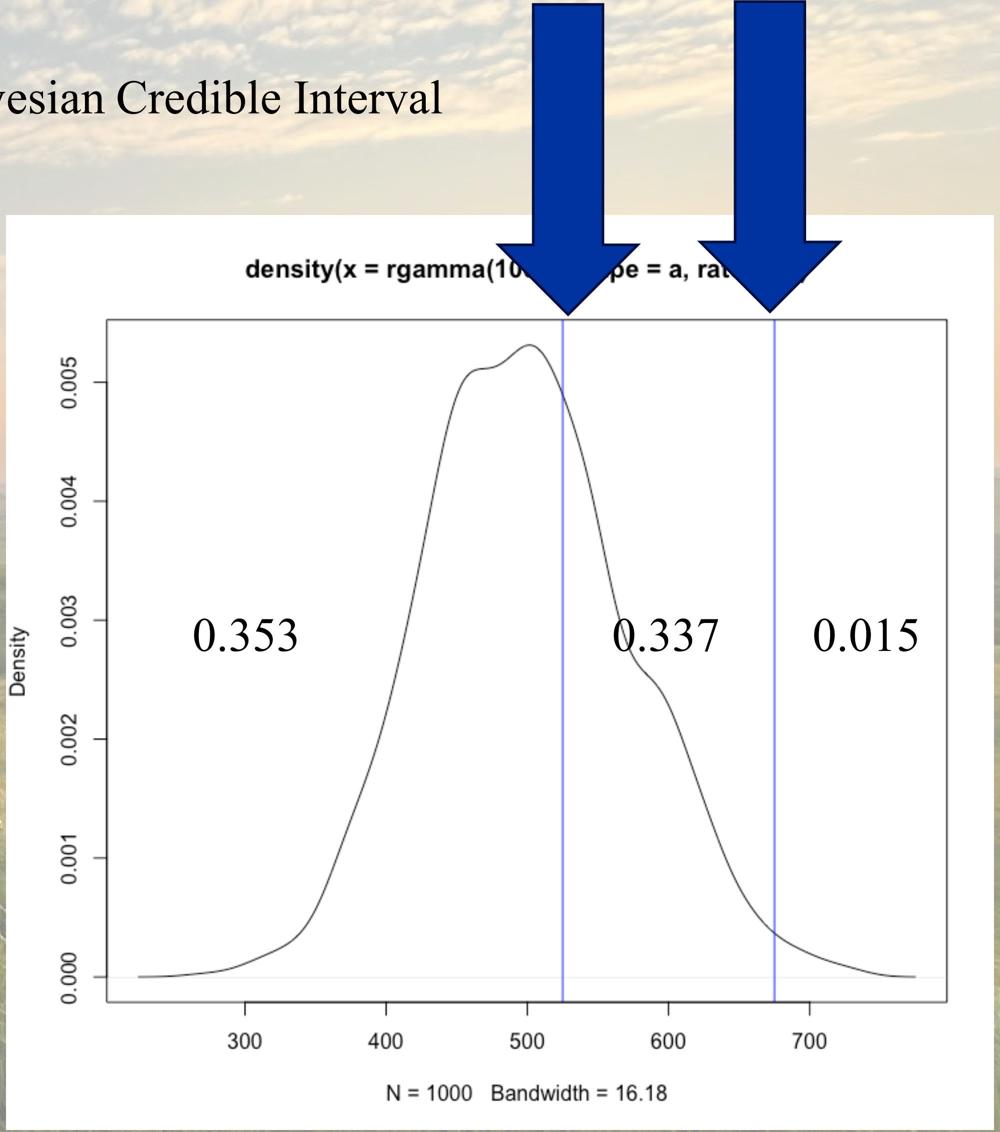
```
plot(density(rgamma(1000, shape = a, rate = B)))
abline(v=600, col = 'red’)
pgamma(q=600,shape = a, rate = B)
1-pgamma(q=600,shape = a, rate = B)
```



# Probability of observing a steer between 525 and 675

Bayesian Credible Interval

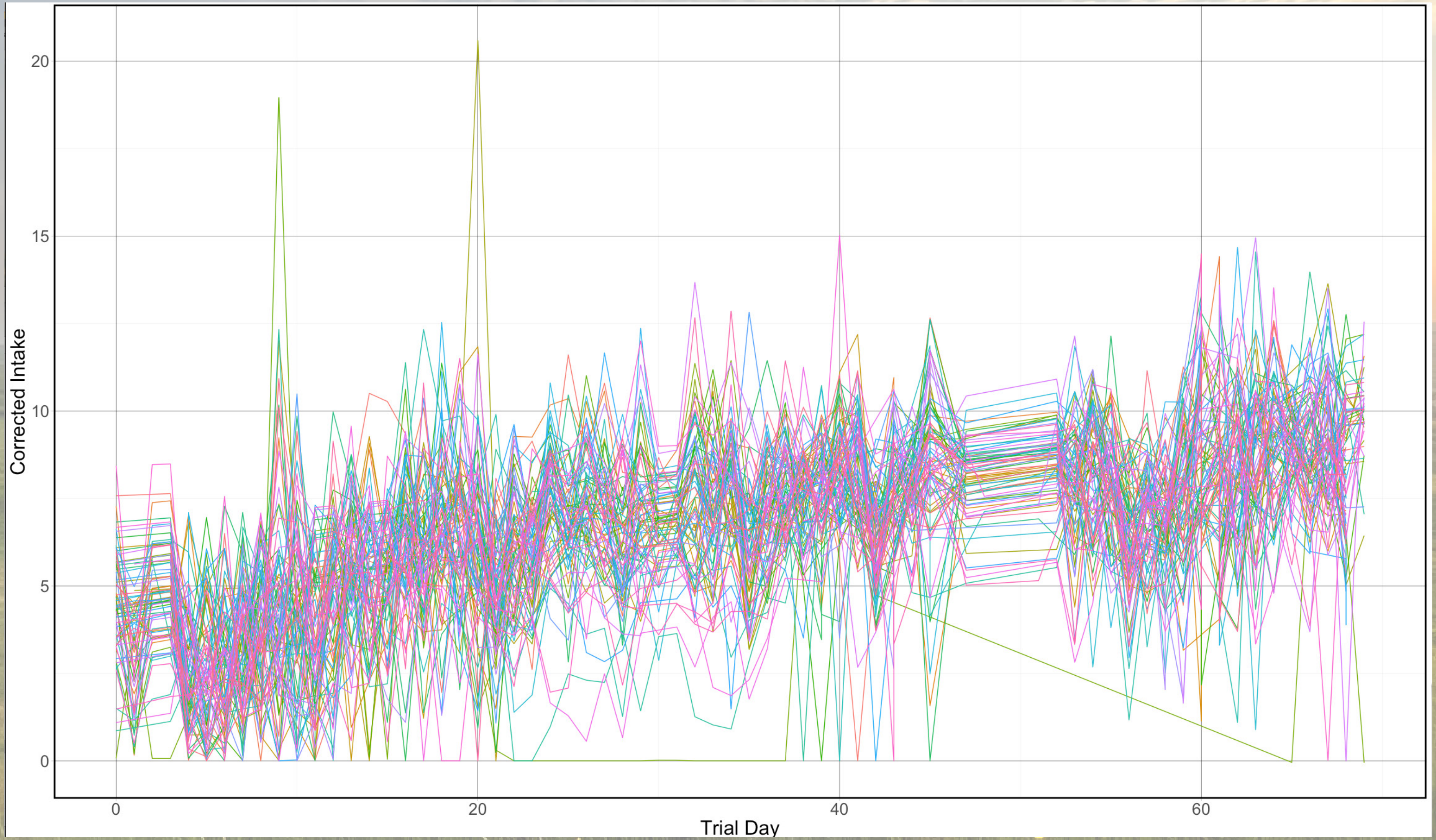
```
plot(density(rgamma(1000, shape = a, rate = B)))
abline(v=525, col = 'blue')
abline(v=675, col = 'blue')
pgamma(q=675,shape = a, rate = B)-
  pgamma(q=525,shape = a, rate = B)
```



# Day-to-day variability in DMI effect on DMI and Gain

LAN trial

# DMI is known to vary from day to day within animal



# Variables

- Dry Matter Intake (DMI)
- Average Dry Matter Intake (uDMI)
- Standard Deviation of Dry Matter Intake (sdDMI)
- Coefficient of Variation of Dry Matter Intake (cvDMI)

# Define hypothesis

- Animals with increased variability in day-to-day feed intake exhibit lower DMI and lower average daily gain

# Find prior knowledge

*Journal of Animal Science*, 2020, Vol. 98, No. 7, 1–9

doi:10.1093/jas/skaa189
Advance Access publication June 26, 2020
**Received:** 3 September 2019 and **Accepted:** 22 June 2020
Ruminant Nutrition

AMERICAN SOCIETY OF ANIMAL SCIENCE

RUMINANT NUTRITION

## Characterization of feeding behavior traits in steers with divergent residual feed intake consuming a high-concentrate diet

Ira L. Parsons, Jocelyn R. Johnson, William C. Kayser, Luis O. Tedeschi, and Gordon E. Carstens[1]

Department of Animal Science, Texas A&M University, College Station, TX 77845

[1]Corresponding author: g-carstens@tamu.edu

ORCiD number: 0000-0003-1883-4911 (L. O. Tedeschi).



## Results and Discussion

### Growth and performance

Performance, feed efficiency, and ultrasound least squared means are presented in Table 3. The initial age of steers at the start of the trials averaged 290 ± 16 d and ranged from 280 to 313 d. The means and SD for ADG and DMI were 1.71 ± 0.26 and 10.1 ± 1.1 kg/d, respectively, which are consistent with growth patterns expected from steers of this breed, weight, and age class. In this study, variation in ADG and mid-test $BW^{0.75}$

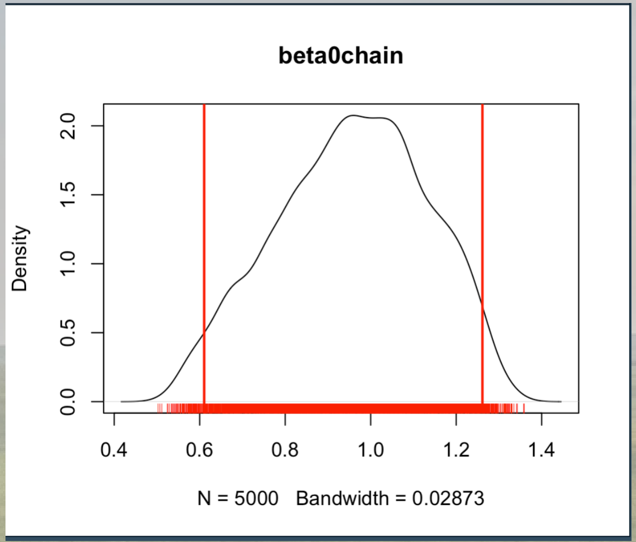
# Factor out model

- Deterministic model: $yi = \beta o + \beta 1xi$
- Conditional model: $[\beta,\sigma^2|y]$ **α** $[y|\beta o, \beta 1, \sigma^2]$
- Factored conditional model: $[\beta,\sigma^2|y]$ **α** $[y|\beta o, \beta 1, \sigma^2][\beta o][\beta 1][\sigma^2]$
- Define posterior distribution model:
  - $[\beta,\sigma^2|y]$ **α** $\prod$ Normal($yi|\ g(\beta o,\beta 1,xi), \sigma^2$)
    X Normal($\beta o|$ 2.29, 0.42)
    X Normal($\beta 1|$ 2.29, 0.42)
    X uniform($\sigma|$ 0, 2)

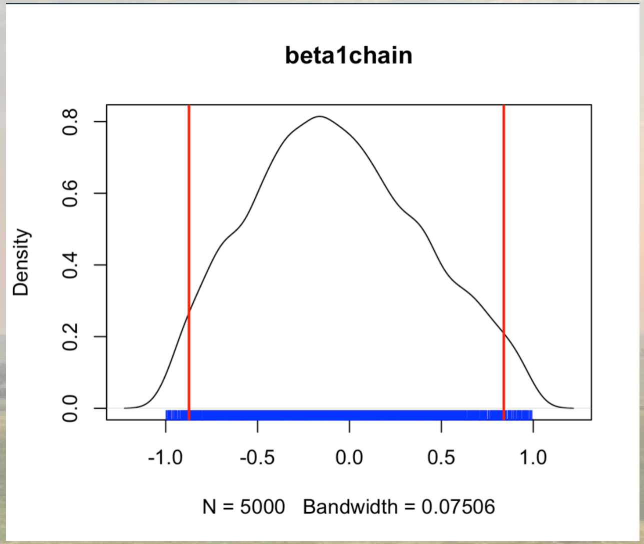
# Define Jags Model

```r

## Jags model ----
# Define the model
model_string <- "
model {
  for (i in 1:N) {
    y[i] ~ dnorm(mu[i], tau)
    mu[i] <- beta0 + beta1 * x[i]
  }

  # Priors
  beta0 ~ dunif(-10, 10)
  beta1 ~ dunif(-1, 1)
  sigma ~ dunif(0, 100)
  tau <- 1 / (sigma * sigma)
}
"

# Data
data_list <- list(
  x = d.dmis$cvDMI,  # Example predictor values
  y = d.lan$Day56_ADG, # Example response values
  N = nrow(d.lan)     # Number of observations
)
```

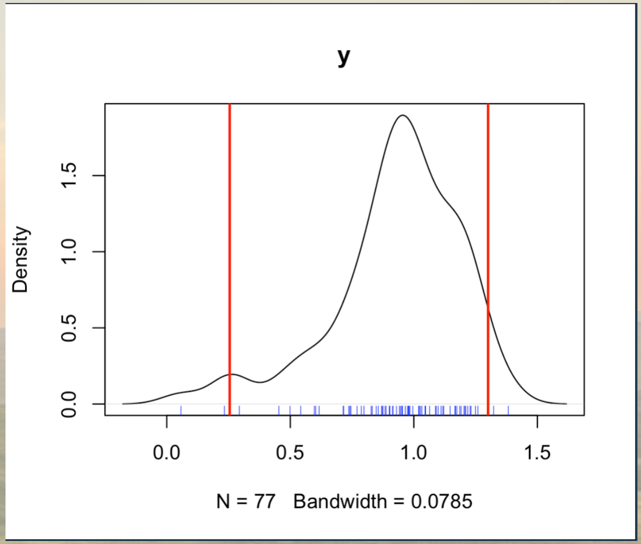
# Results




| 2.5% | 97.5% |
|---|---|
| 0.61 | 1.26 |

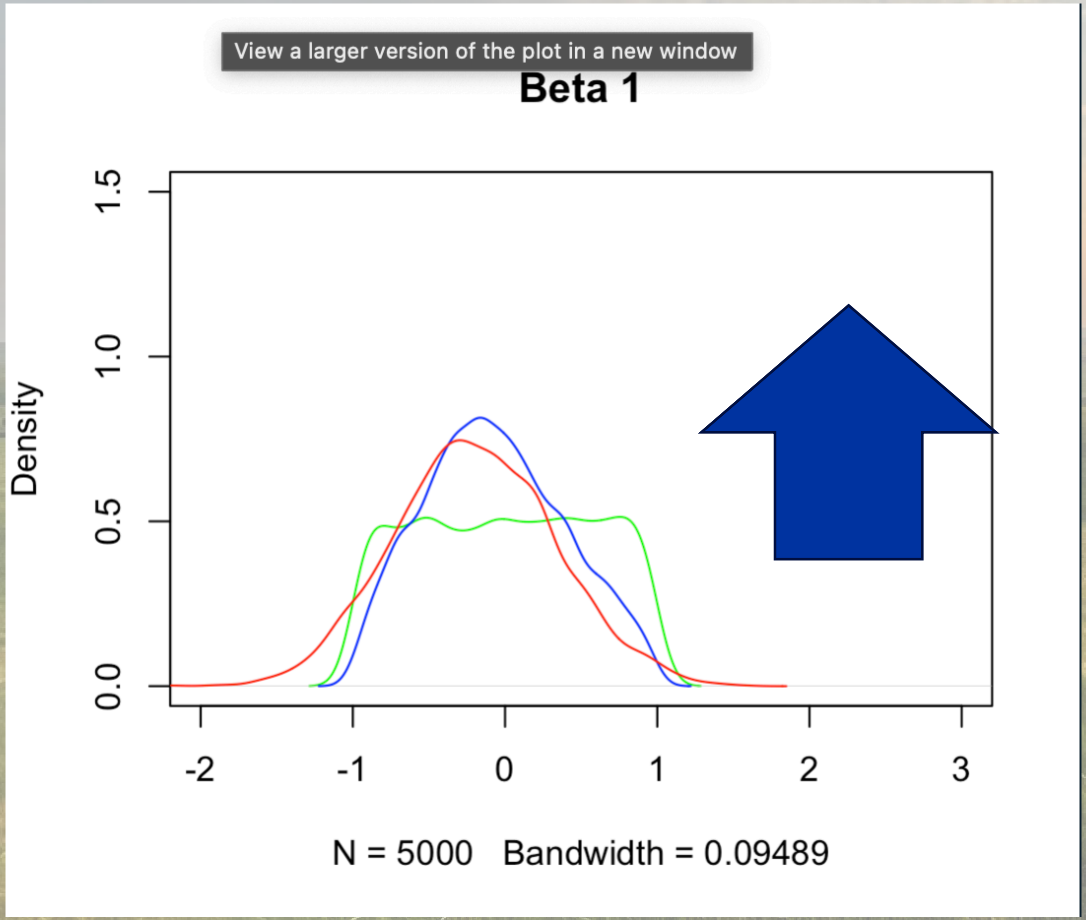



| 2.5% | 97.5% |
|---|---|
| -0.87 | 0.84 |




| 2.5% | 97.5% |
|---|---|
| 0.25 | 1.30 |

# Prior, posterior, and joint distributions




Green – Prior
Blue – Posterior
Red - Joint

# Basics of Bayesian

- Unobserved quantities are random
- Probability is contingent upon the sampling space and definition of the problem
- Joint probabilities are used to quantify likelihood
- Probability distributions are used to describe frequency of data occurring
- Moments are distribution parameters, defining where the data exists in a sampling space
- Choose distributions based upon your data type
- Graph Probability distributions