

Module 4: Bayesian Methods

Lecture 2: Review of Probability

Peter Hoff

Departments of Statistics and Biostatistics
University of Washington

Outline

Beliefs and Probability

Bayes rule

Standard distributions

Posterior summaries

Belief functions

Let F , G , and H be three statements about the world. For example:

$F = \{ \text{a woman will have breast cancer} \}$

$G = \{ \text{a woman will undergo HRT} \}$


$H = \{ \text{a woman has a mutation of BRAC1} \}$

A *belief function* $\text{Be}()$ assigns numbers to statements representing our beliefs:

$$\text{Be}(F|G) \stackrel{?}{>} \text{Be}(F)$$

$$\text{Be}(F|H) \stackrel{?}{>} \text{Be}(F)$$

Axioms of belief

B1 $\text{Be}(\text{not } H|H) \leq \text{Be}(F|H) \leq \text{Be}(H|H)$ 

B2 $\text{Be}(F \text{ or } G|H) \geq \max\{\text{Be}(F|H), \text{Be}(G|H)\}$

B3 $\text{Be}(F \text{ and } G|H)$ can be derived from $\text{Be}(G|H)$ and $\text{Be}(F|G \text{ and } H)$

Do these axioms make sense?

B1: $\text{Be}(F|H)$ is between complete disbelief and complete belief.

B2: Belief in a statement is increasing in the number of possibilities

B3: Deciding on F and G can be done by first deciding on G , then F given G .

Axioms of probability

$F \cup G$ means “ F or G ,” $F \cap G$ means “ F and G ” and \emptyset is the empty set.

P1 $0 = \Pr(\text{not } H|H) \leq \Pr(F|H) \leq \Pr(H|H) = 1$

P2 $\Pr(F \cup G|H) = \Pr(F|H) + \Pr(G|H)$ if $F \cap G = \emptyset$

P3 $\Pr(F \cap G|H) = \Pr(G|H) \Pr(F|G \cap H)$

A probability function, satisfying **P1**, **P2** and **P3**, also satisfies **B1**, **B2** and **B3**.

Examples

[P2] $\Pr(F \cup G|H) = \Pr(F|H) + \Pr(G|H)$ if $F \cap G = \emptyset$

- Genotype is $\{aa, Aa, AA\}$ with probability $\{1/4, 1/2, 1/4\}$.
- $F = \{\text{genotype is } aa\}$, $G = \{\text{genotype is } Aa\}$

$$\begin{aligned}\Pr(\text{one or more } a \text{ alleles}) &= \Pr(F \cup G) \\ &= \Pr(F) + \Pr(G) \\ &= 1/4 + 1/2 = 3/4\end{aligned}$$

[P3] $\Pr(F \cap G|H) = \Pr(G|H) \Pr(F|G \cap H)$

- Parents are AB/AB and AB/ab
- $F = \{\text{child is } A/a \text{ at first locus}\}$, $G = \{\text{child is } B/B \text{ at second locus}\}$

$$\begin{aligned}\Pr(AB/aB) &= \Pr(F \cap G) \\ &= \Pr(G) \times \Pr(F|G) \\ &= \Pr(F|G)/2\end{aligned}$$

Events and partitions

Definition (Partition)

A collection of sets $\{H_1, \dots, H_K\}$ is a partition of another set \mathcal{H} if

1. the events are disjoint, which we write as $H_i \cap H_j = \emptyset$ for $i \neq j$;
2. the union of the sets is \mathcal{H} , which we write as $\bigcup_{k=1}^K H_k = \mathcal{H}$.

If \mathcal{H} is the set of all possible truths and $\{H_1, \dots, H_K\}$ is a partition of \mathcal{H} , then exactly one out of $\{H_1, \dots, H_K\}$ contains the truth.

Examples:

- \mathcal{H} = someone's number of children
 - $\{0, 1, 2, 3 \text{ or more}\}$;
 - $\{0, 1, 2, 3, 4, 5, 6, \dots\}$.
- \mathcal{H} = the relationship between a genotype and heart disease
 - $\{\text{some relationship, no relationship}\}$;
 - $\{\text{negative correlation, zero correlation, positive correlation}\}$.


Bayes' rule

The axioms of probability imply the following:

Rule of total probability : $\sum_{k=1}^K \Pr(H_k) = 1$


Rule of marginal probability : $\Pr(E) = \sum_{k=1}^K \Pr(E \cap H_k)$

$= \sum_{k=1}^K \Pr(E|H_k) \Pr(H_k)$



Bayes' rule : $\Pr(H_j|E) = \frac{\Pr(E|H_j) \Pr(H_j)}{\Pr(E)}$

$= \frac{\Pr(E|H_j) \Pr(H_j)}{\sum_{k=1}^K \Pr(E|H_k) \Pr(H_k)}$



An example

A sonogram indicates someone will have twin girls.

What is the probability that the girls are monozygotic?

- $\mathcal{H} = \{H_1, H_2\} = \{ \text{monozygotic} , \text{dizygotic} \}$
- $E = \{ \text{twins are girls} \}$

An example

- $\mathcal{H} = \{H_1, H_2\} = \{ \text{monozygotic} , \text{dizygotic} \}$
- $E = \{ \text{twins are girls} \}$

Observed data: twins are girls.

Prior information: One third of twins are monozygotic (identical).

$$\begin{aligned}\Pr(H_1|E) &= \frac{\Pr(E|H_1) \Pr(H_1)}{\Pr(E)} \\ &= \frac{\Pr(E|H_1) \Pr(H_1)}{\Pr(E|H_1) \Pr(H_1) + \Pr(E|H_2) \Pr(H_2)} \\ &= \frac{1/2 \times 1/3}{1/2 \times 1/3 + 1/4 \times 2/3} \\ &= \frac{1/6}{1/6 + 1/6} \\ &= \frac{1}{2} > \frac{1}{3} = \Pr(H_1)\end{aligned}$$

Bayesian inference

$\{H_1, \dots, H_K\}$ often refer to disjoint hypotheses or states of nature

E refers to the outcome of a survey, study or experiment.

Post-experimental evaluation of hypotheses are via the posterior odds ratio:

$$\begin{aligned}
 \frac{\Pr(H_i|E)}{\Pr(H_j|E)} &= \frac{\Pr(E|H_i) \Pr(H_i) / \Pr(E)}{\Pr(E|H_j) \Pr(H_j) / \Pr(E)} \\
 &= \frac{\Pr(E|H_i) \Pr(H_i)}{\Pr(E|H_j) \Pr(H_j)} \\
 &= \frac{\Pr(E|H_i)}{\Pr(E|H_j)} \times \frac{\Pr(H_i)}{\Pr(H_j)} \\
 &= \text{"Likelihood ratio"} \times \text{"prior odds"}
 \end{aligned}$$

Twin example

Prior odds:

$$\frac{\Pr(H_1)}{\Pr(H_2)} = \frac{1/3}{2/3} = 1/2$$

Prior favors H_2

Likelihood ratio:

$$\frac{\Pr(E|H_1)}{\Pr(E|H_2)} = \frac{1/2}{1/4} = 2$$

Data is better explained by H_1

Posterior odds:

$$\frac{\Pr(H_1)}{\Pr(H_2)} \frac{\Pr(E|H_1)}{\Pr(E|H_2)} = 1$$

Independence

F and G are conditionally independent given H if

$$\Pr(F \cap G|H) = \Pr(F|H) \Pr(G|H)$$

Interpretation: By Axiom **P3**, the following is always true:

$$\Pr(F \cap G|H) = \Pr(G|H) \Pr(F|H \cap G).$$

If F and G are conditionally independent given H , then we must have

$$\begin{aligned} \Pr(G|H) \Pr(F|H \cap G) &\stackrel{\text{always}}{=} \Pr(F \cap G|H) &\stackrel{\text{independence}}{=} \Pr(F|H) \Pr(G|H) \\ \Pr(G|H) \Pr(F|H \cap G) &= \Pr(F|H) \Pr(G|H) \\ \Pr(F|H \cap G) &= \Pr(F|H). \end{aligned}$$

Conditional independence therefore implies that $\Pr(F|H \cap G) = \Pr(F|H)$.

If H is known and F and G are conditionally independent given H , then knowing G does not change our belief about F .

Independence

Example:

$F = \{ \text{a patient will develop cancer} \}$

$G = \{ \text{the parents' genotypes} \}$

$H = \{ \text{a patient's genotype} \}$

$$\Pr(F|H) \stackrel{?}{=} \Pr(F|G, H)$$



Discrete random variables

Let Y be a *random variable*, an unknown numerical quantity.

Let \mathcal{Y} be the set of all possible values of Y .

Y is *discrete* if the set of possible outcomes is *countable*, meaning that \mathcal{Y} can be expressed as $\mathcal{Y} = \{y_1, y_2, \dots\}$.

Examples


- Y = number of people in a population with a specific allele
- Y = number of children of a randomly sampled person
- Y = number of years of education of a randomly sampled person

Probability distributions and densities

$\Pr(Y = y)$ is the probability that the outcome Y takes on the value y .

PDF: $\Pr(Y = y) = p(y)$ is often called the *probability density function* (pdf) of Y

1. $0 \leq p(y) \leq 1$ for all $y \in \mathcal{Y}$;

2. $\sum_{y \in \mathcal{Y}} p(y) = 1$. 

We can derive various probabilities from $p(y)$:

$$\Pr(Y \in A) = \sum_{y \in A} p(y)$$

If A and B are disjoint subsets of \mathcal{Y} , then

$$\begin{aligned} \Pr(Y \in A \text{ or } Y \in B) &\equiv \Pr(Y \in A \cup B) &= \Pr(Y \in A) + \Pr(Y \in B) \\ &= \sum_{y \in A} p(y) + \sum_{y \in B} p(y). \end{aligned}$$

Probability densities

If (to a rough approximation) $\mathcal{Y} = \mathbb{R}$, then we cannot define $\Pr(Y \leq 5)$ as equal to $\sum_{y \leq 5} p(y)$ because the sum does not make sense.

Instead, we define a probability density as a function $p(y)$ such that

$$\Pr(Y \in A) = \int_A p(y) dy$$

1. $0 \leq p(y)$ for all $y \in \mathcal{Y}$;
2. $\int_{\mathbb{R}} p(y) dy = 1$.

As in the discrete case, probability statements about Y can be derived from the pdf: $\Pr(Y \in A) = \int_{y \in A} p(y) dy$, and if A and B are disjoint subsets of \mathcal{Y} , then

$$\begin{aligned} \Pr(Y \in A \text{ or } Y \in B) &\equiv \Pr(Y \in A \cup B) &= &\Pr(Y \in A) + \Pr(Y \in B) \\ &= &\int_{y \in A} p(y) dy + \int_{y \in B} p(y) dy. \end{aligned}$$

Notes: Unlike the discrete case,

- $p(y)$ can be larger than 1;
- $p(y)$ is not “the probability that $Y = y$.”

However, if $p(y_1) > p(y_2)$ we will sometimes informally say that y_1 “has a higher probability” than y_2 .

The binary distribution

Let $\mathcal{Y} = \{0, 1\}$.

The outcome Y has a *binary distribution with probability θ* if

$$\Pr(Y = y|\theta) = p(y|\theta) = \begin{cases} \theta & \text{if } y = 1 \\ (1 - \theta) & \text{if } y = 0 \end{cases}$$

Alternatively, we can write

$$\Pr(Y = y|\theta) = p(y|\theta) = \theta^y(1 - \theta)^{1-y}$$



Independent binary outcomes

Suppose the prevalence of an allele in a population is θ .

Let Y_1, \dots, Y_n indicate the presence of the allele for n individuals randomly sampled from the population.

$$\begin{aligned}\Pr(Y_1 = y_1, \dots, Y_n = y_n | \theta) &= p(y_1, \dots, y_n | \theta) \\ &= \left(\theta^{y_1} (1 - \theta)^{1 - y_1} \right) \times \dots \times \left(\theta^{y_n} (1 - \theta)^{1 - y_n} \right) \\ &= \theta^{\sum y_i} (1 - \theta)^{n - \sum y_i} \quad \text{💬}\end{aligned}$$

Note that $p(y_1, \dots, y_n | \theta)$ depends only on $\sum y_i$.

Often, we only record $y = \sum y_i$ and n .

The binomial distribution

What is the probability that y people in a sample of size n will have the allele?

Consider all n -sequences with y 1's:

$$\begin{array}{rcl} \Pr(Y_1 = 0, Y_2 = 1, Y_3 = 0, \dots, Y_n = 1 | \theta) & = & \theta^y (1 - \theta)^{n-y} \\ & \vdots & \\ \Pr(Y_1 = 1, Y_2 = 0, Y_3 = 1, \dots, Y_n = 0 | \theta) & = & \theta^y (1 - \theta)^{n-y} \end{array}$$

There are $\binom{n}{y}$ such sequences, so

$$\Pr(\sum Y_i = y | \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$



The binomial distribution

Let $\mathcal{Y} = \{0, 1, 2, \dots, n\}$ for some positive integer n . The outcome $Y \in \mathcal{Y}$ has a *binomial distribution with probability θ* if

$$\Pr(Y = y|\theta) = \text{dbinom}(y, n, \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}.$$

For example, if $\theta = .25$ and $n = 4$, we have:

$$\Pr(Y = 0|\theta = .25) = \binom{4}{0} (.25)^0 (.75)^4 = .316$$

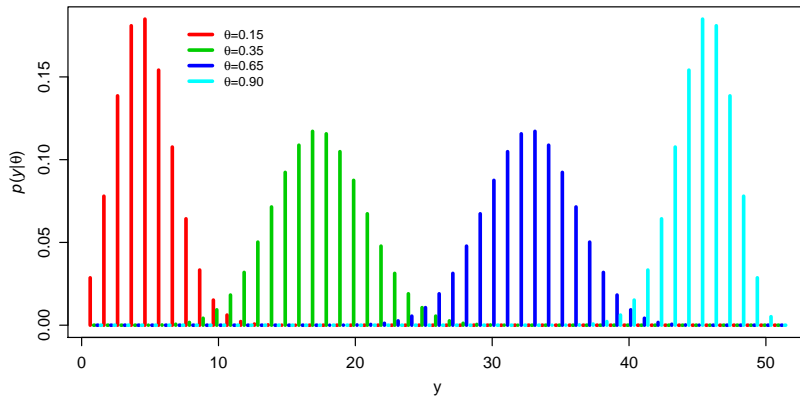
$$\Pr(Y = 1|\theta = .25) = \binom{4}{1} (.25)^1 (.75)^3 = .422$$

$$\Pr(Y = 2|\theta = .25) = \binom{4}{2} (.25)^2 (.75)^2 = .211$$

$$\Pr(Y = 3|\theta = .25) = \binom{4}{3} (.25)^3 (.75)^1 = .047$$

$$\Pr(Y = 4|\theta = .25) = \binom{4}{4} (.25)^4 (.75)^0 = .004.$$

The binomial distribution



Likelihood inference for θ

$y = 9$ out of $n = 50$ randomly sampled subjects have a particular allele.

What is the allele frequency in the general population?

The likelihood:

$$\Pr(Y = 9|\theta = 0.15) = 0.033$$

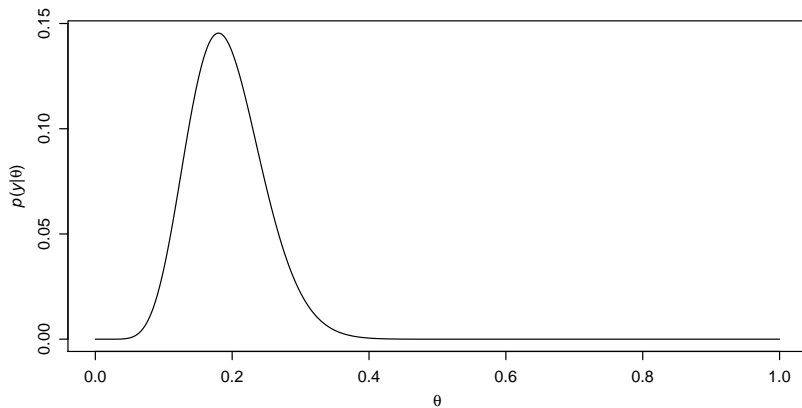
$$\Pr(Y = 9|\theta = 0.35) = 0.004$$

$$\Pr(Y = 9|\theta = 0.65) \approx 10^{-11}$$

$$\Pr(Y = 9|\theta = 0.90) \approx 10^{-32}$$

As a function of θ , $\Pr(Y = y|\theta) = p(y|\theta)$ is called the *likelihood of θ* .

The binomial likelihood



Posterior inference via Bayes' rule

Suppose

- $\Theta = \{0.01, 0.02, 0.03, \dots, 0.99, 1.00\}$ and
- $p(\theta) = 1/100$ for each $\theta \in \Theta$.

Bayes' rule says

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\sum_{\theta \in \Theta} p(y|\theta)p(\theta)}$$

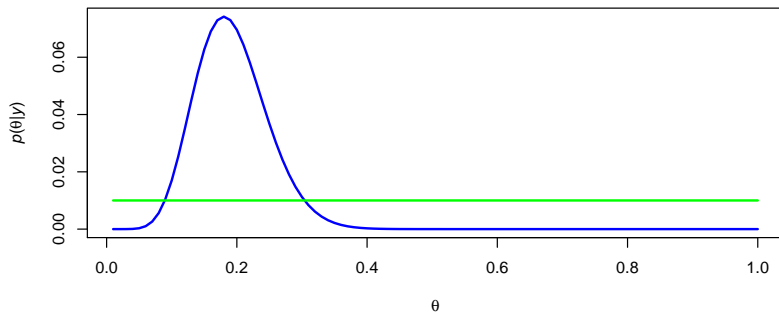
```
n<-50
y<-9

theta<-seq(.01,1,length=100)

py.theta<-dbinom(y,n,theta)
ptheta<-rep(1/100,length=length(theta))

ptheta.y<- py.theta*ptheta / sum( py.theta*ptheta )
```

The posterior distribution



The posterior distribution

Now suppose

- $\Theta = [0, 1]$ and
- $p(\theta) = 1$ for each $\theta \in \Theta$.

Bayes' rule says

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_0^1 p(y|\theta)p(\theta)}$$

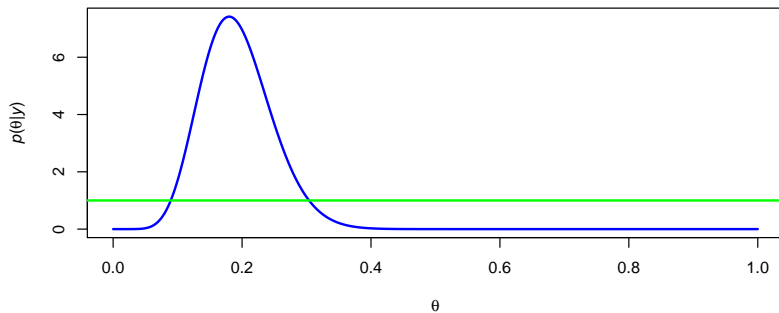
The next lecture shows $p(\theta|y)$ is a *beta* $(1 + y, 1 + n - y)$ *distribution*, and

$$\mathbb{E}[\theta|y] = \frac{1}{2} \left(\frac{2}{2+n} \right) + \frac{y}{n} \left(\frac{n}{2+n} \right)$$

```
n<-50
y<-9

theta<-seq(0,1,length=500)
ptheta.y<-dbeta(theta,1+y,1+n-y)
```

The posterior distribution



The Poisson distribution

Let $\mathcal{Y} = \{0, 1, 2, \dots\}$. The outcome $Y \in \mathcal{Y}$ has a *Poisson distribution* with mean θ if

$$\Pr(Y = y|\theta) = \text{dpois}(y, \theta) = \theta^y e^{-\theta} / y!.$$

For example, if $\theta = 2.1$ (the 2006 U.S. fertility rate),

$$\Pr(Y = 0|\theta = 2.1) = (2.1)^0 e^{-2.1} / (0!) = .12$$

$$\Pr(Y = 1|\theta = 2.1) = (2.1)^1 e^{-2.1} / (1!) = .26$$

$$\Pr(Y = 2|\theta = 2.1) = (2.1)^2 e^{-2.1} / (2!) = .27$$

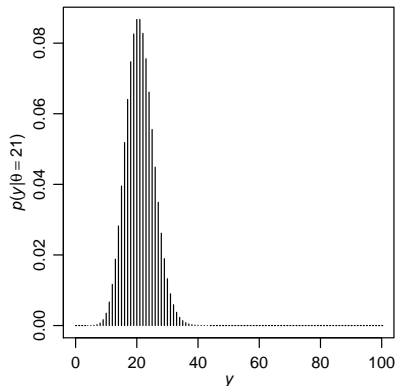
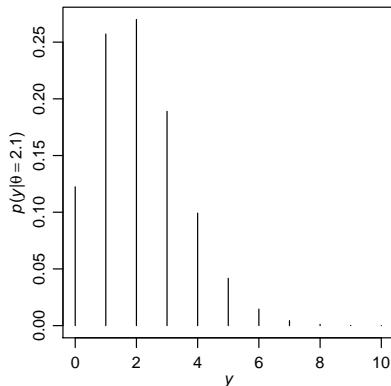
$$\Pr(Y = 3|\theta = 2.1) = (2.1)^3 e^{-2.1} / (3!) = .19$$

$$\vdots$$

$$\vdots$$

$$\vdots$$

The Poisson distribution



Poisson distributions with means of 2.1 and 21.

The Poisson likelihood

Let Y_i be the number of intestinal tumors for mouse i , $i = 1, \dots, n$.

What is the mean tumor count in this population?

The likelihood:

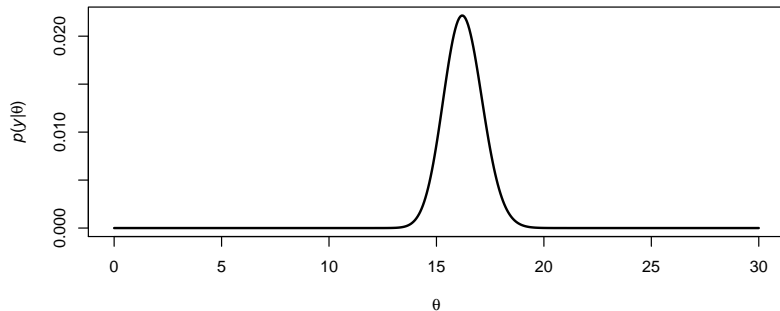
$$\begin{aligned}
 \Pr(Y_1 = y_1, \dots, Y_n = y_n | \theta) &= p(y_1, \dots, y_n | \theta) \\
 &= \prod_{i=1}^n p(y_i | \theta) \\
 &= \prod_{i=1}^n \theta^{y_i} e^{-\theta} / y_i! \\
 &= \theta^{\sum y_i} e^{-n\theta} \times (\prod y_i!)^{-1}
 \end{aligned}$$

Simplification: Let $Y = \sum Y_i$. Then $\{Y | \theta\} \sim \text{Poisson}(n\theta)$ and so

$$\Pr(Y = y | \theta) = \theta^y e^{-n\theta} \times (n^y / y!)$$

The Poisson likelihood

Suppose $n = 20$ and $y = \sum y_i = 324$ ($y/n = 16.2$).

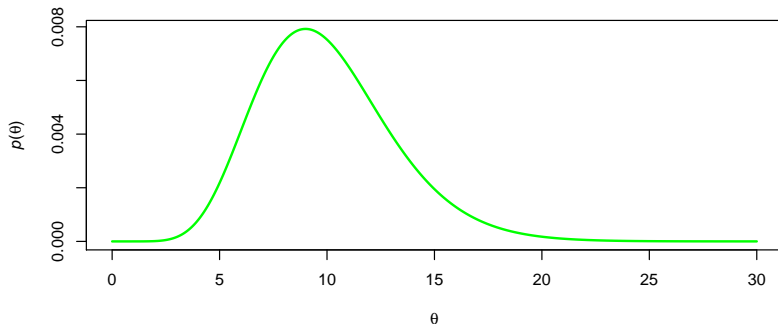


Posterior inference

Similar populations of mice suggest $\theta \approx 10$.

A suitable prior distribution for θ would reflect this:

$$\begin{aligned}\theta &\sim \text{gamma}(10, 1) \\ E[\theta] &= 10 \\ \text{SD}[\theta] &= \sqrt{10} \approx 3.16\end{aligned}$$



Discrete approximation to the posterior

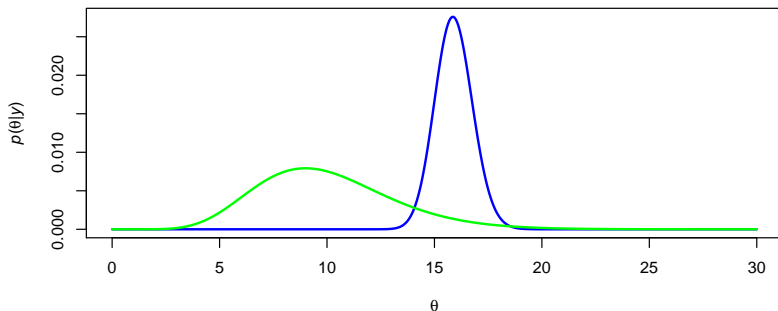
```

n<-20
sy<-324

theta<-seq(0,30,length=500)

psy.theta<-dpois(sy, n*theta)
p.theta<-dgamma(theta,10,1) ; p.theta<-p.theta/sum(p.theta)
p.theta.y<- psy.theta * p.theta / sum( psy.theta * p.theta)

```



The gamma posterior distribution

It can be shown that if

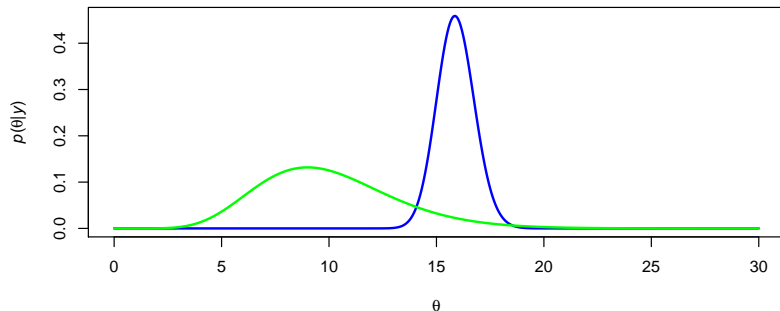
- $\theta \sim \text{gamma}(a, b)$ and
- $Y_1, \dots, Y_n | \theta \sim \text{Poisson}(\theta)$ then

$$\{\theta | \sum Y_i = y\} \sim \text{gamma}(a + y, b + n)$$

$$\begin{aligned} \mathbb{E}[\theta | \sum Y_i = y] &= \frac{a + y}{b + n} \\ &= \frac{a}{b} \left(\frac{b}{b + n} \right) + \frac{y}{n} \left(\frac{n}{b + n} \right) \end{aligned}$$

The gamma posterior distribution

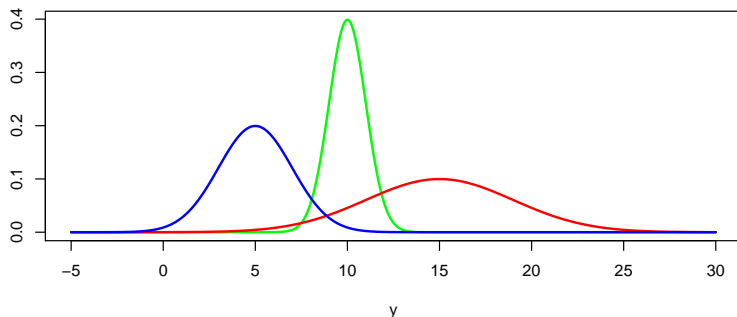
```
a<-10 ; b<-1                                # prior
sy<-324 ; n<-20                             # data
theta<-seq(0,30,length=500)
p.theta.y<- dgamma(theta, a+sy, b+n)        # posterior
```



The normal distribution

Let $\mathcal{Y} = (-\infty, \infty)$. The outcome $Y \in \mathcal{Y}$ has a *normal distribution* with mean θ and variance σ^2 if

$$p(y|\theta, \sigma^2) = \text{dnorm}(y, \theta, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{y - \theta}{\sigma} \right)^2 \right\}.$$



The normal posterior distribution

It can be shown that if

- $\theta \sim \text{normal}(\mu_0, \tau_0^2)$ and
- $Y_1, \dots, Y_n | \theta \sim \text{normal}(\theta, \sigma^2)$

then

$$\{\theta | y_1, \dots, y_n\} \sim \text{normal}(\mu_n, \tau_n^2)$$

where

$$\begin{aligned}\text{Var}[\theta | y_1, \dots, y_n] = \tau_n^2 &= [1/\tau_0^2 + n/\sigma^2]^{-1} \\ 1/\tau_n^2 &= 1/\tau_0^2 + n/\sigma^2\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[\theta | y_1, \dots, y_n] = \mu_n &= \frac{\mu_0/\tau_0^2 + \bar{y}n/\sigma^2}{1/\tau_0^2 + n/\sigma^2} \\ &= \mu_0 \left(\frac{1/\tau_0^2}{1/\tau_0^2 + n/\sigma^2} \right) + \bar{y} \left(\frac{n/\sigma^2}{1/\tau_0^2 + n/\sigma^2} \right)\end{aligned}$$

Describing posterior location

The posterior *mean* or *expectation* of an unknown quantity θ is given by

$$E[\theta|y] = \int_{\theta \in \Theta} \theta p(\theta|y) d\theta$$

The mean is the center of mass of the distribution.

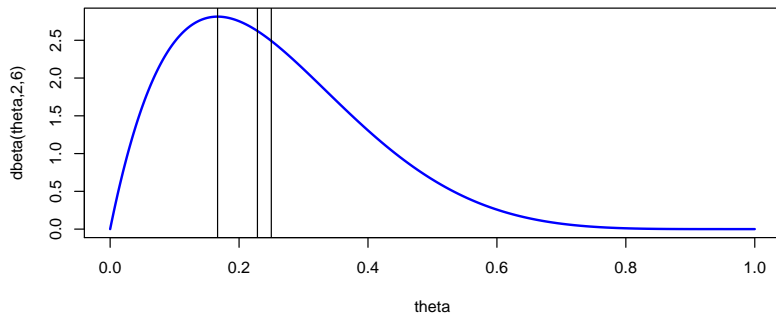
However, it is not in general equal to either of

the *mode*: “the most probable value of θ ,” or

the *median*: “the value of θ in the middle of the distribution.”

For skewed distributions the mean can be far from a “typical” sample value

Mean, median and mode



Variance and SD

How spread out is $p(\theta|y)$?

Popular measures of spread are the *variance* and *standard deviation* :

$$\begin{aligned}\text{Var}[\theta] &= \text{E}[(\theta - \text{E}[\theta|y])^2] \\ &= \text{E}[\theta^2|y] - \text{E}[\theta|y]^2.\end{aligned}$$

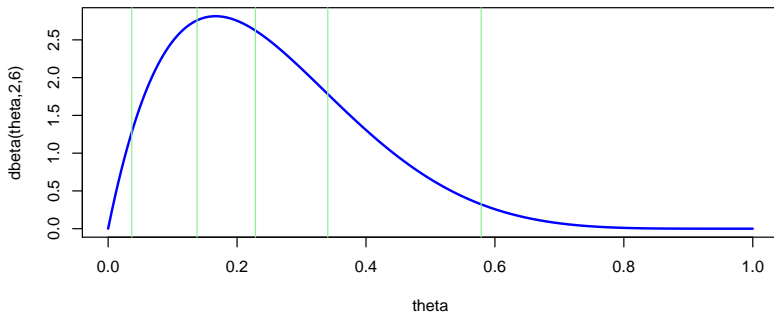
The variance is the average squared distance of θ to $\text{E}[\theta|y]$.

The *standard deviation* is the square root of the variance, and has the same units as θ .

Quantiles

Alternative measures of spread are based on *quantiles*.

The α -quantile is the value θ_α such that $\Pr(\theta \leq \theta_\alpha | y) = \alpha$.

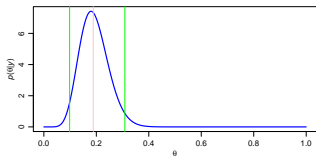


2.5, 25, 50 75 and 97.5 quantiles of a beta distribution

Quantile-based confidence intervals

A popular way to construct a confidence interval is with quantiles:

Allele frequency example: $\theta|y \sim \text{beta}(1+y, 1+n-y)$

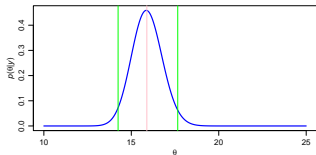


```
a<-1 ; b<-1      #prior
n<-50 ; y<-9      #data

qbeta(c(.025, .5, .975), a+y, b+n-y)

## [1] 0.09824 0.18834 0.30873
```

Tumor count example: $\theta|y \sim \text{gamma}(a+y, b+n)$



```
a<-10 ; b<-1      #prior
y<-324 ; n<-20     #data

qgamma(c(.025, .5, .975), a+y, b+n)

## [1] 14.24 15.89 17.66
```

Monte Carlo approximation

Posterior descriptions involve integrals we'd often like to avoid:


$$E[\theta|y] = \int \theta p(\theta|y) d\theta$$

$$\text{median}[\theta|y] = \theta_{1/2} : \int_{-\infty}^{\theta_{1/2}} p(\theta|y) = 1/2$$

$$\Pr(\theta_1 < \theta_2 | y_1, y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\theta_2} p(\theta_1, \theta_2 | y_1, y_2) d\theta_1 d\theta_2$$

We can easily approximate such integrals arbitrarily closely with **Monte Carlo approximation**.

Monte Carlo approximation

The basic principle: If $\theta^{(1)}, \dots, \theta^{(S)}$  iid $p(\theta|y)$, then

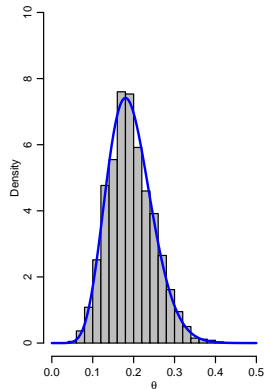
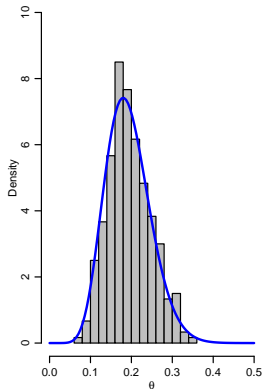
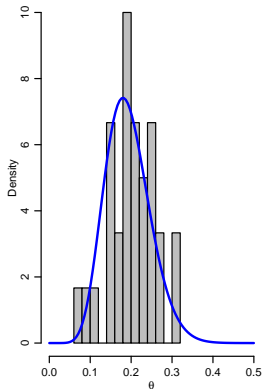
$$\text{histogram}\{\theta^{(1)}, \dots, \theta^{(S)}\} \approx p(\theta|y)$$

This implies that

$$\begin{aligned} \frac{1}{S} \sum \theta^{(s)} &\approx E[\theta|y] \\ \frac{\#\{\theta^{(s)} < c\}}{S} &\approx \Pr(\theta < c|y) \\ \text{median}\{\theta^{(1)}, \dots, \theta^{(S)}\} &\approx \text{median}[\theta|y] \end{aligned}$$

etc. with the approximation improving with increasing S .

Monte Carlo approximation



Monte Carlo approximation

```
## correct values
qbeta(c(.025, .5, .975), a+y, b+n-y)

## [1] 0.09824 0.18834 0.30873

## Monte Carlo simulated data
theta.sim<-rbeta(3000, a+y, b+n-y)

## MC estimate based on 30 sims
quantile(theta.sim[1:30], c(.025, .5, .975))

##      2.5%      50%      97.5%
## 0.1273 0.1763 0.2831

## MC estimates based on 300 sims
quantile(theta.sim[1:300], c(.025, .5, .975))

##      2.5%      50%      97.5%
## 0.1019 0.1901 0.3313

## MC estimates based on 3000 sims
quantile(theta.sim[1:3000], c(.025, .5, .975))

##      2.5%      50%      97.5%
## 0.1002 0.1866 0.3164
```

Monte Carlo for two sample comparisons

Suppose we want to compare the prevalence of a gene in two populations.

- θ_1 = prevalence rate in population 1
- θ_2 = prevalence rate in population 2

Prior:

$$\theta_1, \theta_2 \sim \text{uniform}(0, 1)$$

Data:

- $n_1 = 50, y_1 = 10, y/n = 0.200$
- $n_2 = 27, y_2 = 8, y/n = 0.296$

Posterior:

- $\theta_1 \sim \text{beta}(1 + y_1, 1 + n_1 - y_1) = \text{beta}(11, 41)$
- $\theta_2 \sim \text{beta}(1 + y_2, 1 + n_2 - y_2) = \text{beta}(9, 20)$

Monte Carlo for two sample comparisons

```

a<-1 ; b<-1
n1<-50 ; y1<-10
n2<-27 ; y2<-8

theta1.sim<-rbeta(5000,a+y1,b+n1-y1)
theta2.sim<-rbeta(5000,a+y2,b+n2-y2)

mean(theta1.sim<theta2.sim)

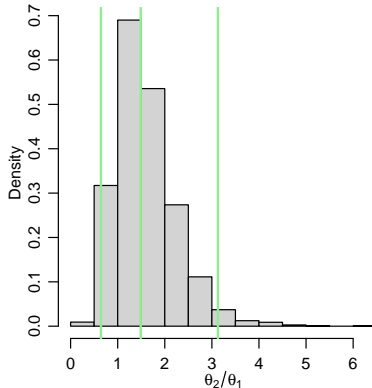
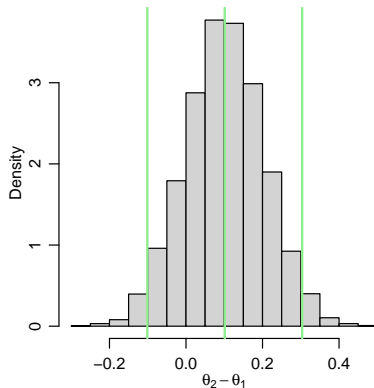
## [1] 0.8368

quantile( theta1.sim-theta2.sim,prob=c(.025,.5,.975))

##      2.5%      50%      97.5%
## -0.3036 -0.1014  0.1011

```

Monte Carlo for two sample comparisons



Summary

- Probability distributions encapsulate information
 - $p(\theta)$ describes prior information
 - $p(y|\theta)$ describes information about y for each θ
 - $p(\theta|y)$ describes posterior information
- Posterior distributions can be calculated via Bayes' rule

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta) d\theta}$$

- Posteriors can be plotted/summarized with commands in R
 - dbeta, dgamma, dnorm
 - qbeta, qgamma, qnorm
- Posterior integrals can be approximated with Monte Carlo methods.
 1. simulate $\theta^{(1)}, \dots, \theta^{(S)} \sim \text{iid } p(\theta|y)$
 2. approximate integrals with simulation averages.