STATS 551

Single-parameter Models

Yang Chen

Department of Statistics

University of Michigan

ychenang@umich.edu

January 27, 2020

Overview

Plan

Estimating a probability from binomial data

- Data: a sequence of 'Bernoulli trials', either 0 or 1.

Estimating a probability from binomial data

- Data: a sequence of 'Bernoulli trials', either 0 or 1.
- Binomial sampling model

$$p(y|\theta) = \left(\begin{array}{c} n \\ y \end{array} \right) \theta^y (1-\theta)^{n-y}.$$

Estimating a probability from binomial data

- Data: a sequence of 'Bernoulli trials', either 0 or 1.
- Binomial sampling model

$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}.$$

- Example: estimating probability of female birth.

Estimating a probability from binomial data

- Data: a sequence of 'Bernoulli trials', either 0 or 1.
- Binomial sampling model

$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}.$$

- Example: estimating probability of female birth.
- Prior for θ: uniform on $(0, 1)$.

Estimating a probability from binomial data

- Data: a sequence of 'Bernoulli trials', either 0 or 1.
- Binomial sampling model

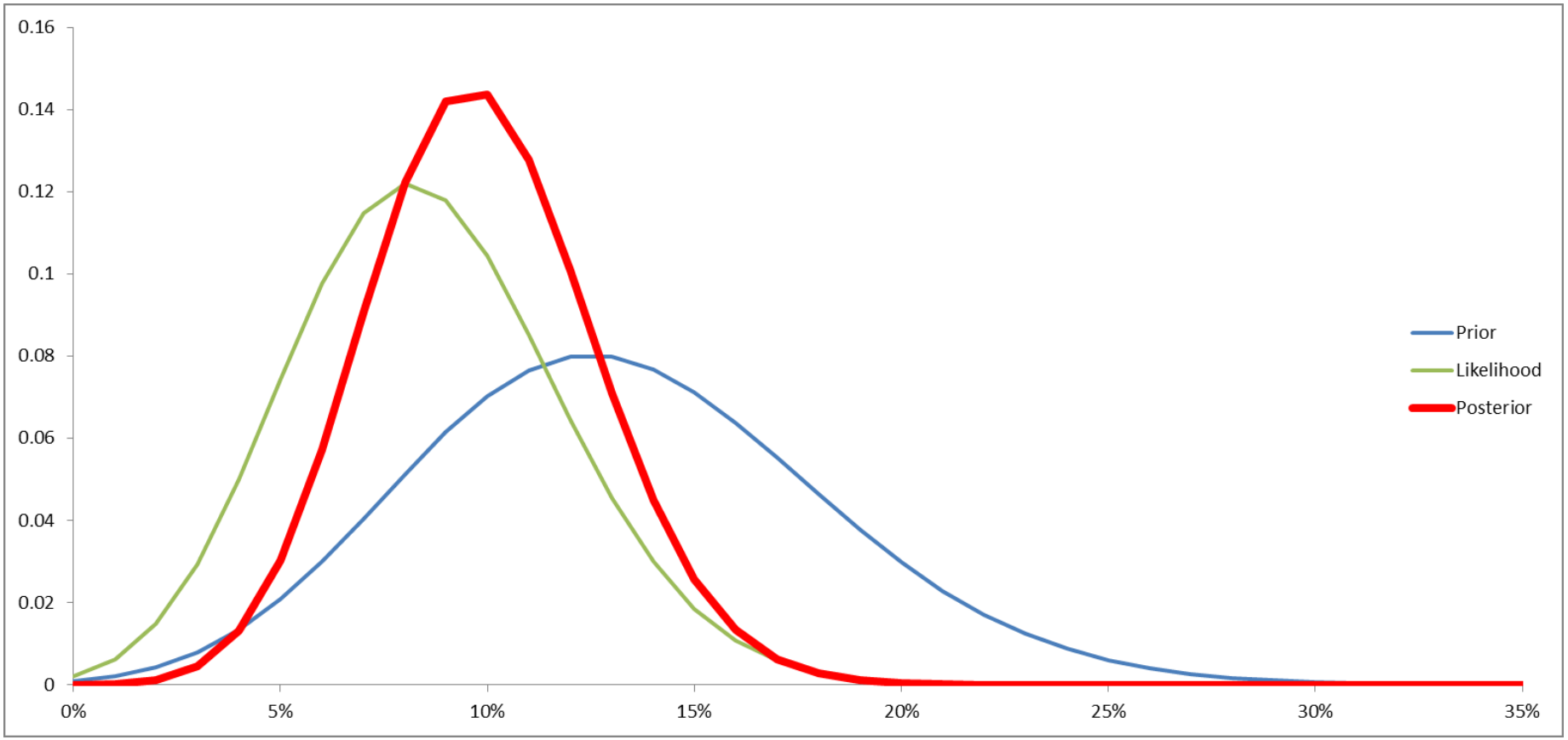
$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}.$$

- Example: estimating probability of female birth.
- Prior for θ: uniform on $(0, 1)$.
- Posterior for θ:

$$\theta|y \sim \text{Beta}(y+1, n-y+1).$$

Plan

Prior and Posterior



Summarizing posterior inference

- Output from Bayesian inference:

  posterior distribution $p(\theta|y)$.

Summarizing posterior inference

- Output from Bayesian inference:

$$\text{posterior distribution } p(\theta|y).$$

- Numerical summaries:
  1. mean
  2. median
  3. mode
  4. standard deviation
  5. interquartile range

Summarizing posterior inference

- Output from Bayesian inference:

$$\text{posterior distribution } p(\theta|y).$$

- Numerical summaries:
  1. mean
  2. median
  3. mode
  4. standard deviation
  5. interquartile range

- Posterior interval

Summarizing posterior inference: Binomial Example

- Output from Bayesian inference:

  posterior distribution $p(\theta|y) = \text{Beta}(y+1, n-y+1)$.

Summarizing posterior inference: Binomial Example

- Output from Bayesian inference:

  posterior distribution $p(\theta|y) = \text{Beta}(y+1, n-y+1)$.

- Numerical summaries:
  1. mean

Summarizing posterior inference: Binomial Example

- Output from Bayesian inference:

$$\text{posterior distribution } p(\theta|y) = \text{Beta}(y+1, n-y+1).$$

- Numerical summaries:
  1. mean $= \frac{y+1}{n+2}$
  2. median

Summarizing posterior inference: Binomial Example

- Output from Bayesian inference:

$$\text{posterior distribution } p(\theta|y) = \text{Beta}(y+1, n-y+1).$$

- Numerical summaries:
  1. mean $= \frac{y+1}{n+2}$
  2. median $\approx \frac{y+2/3}{n+4/3}$
  3. mode

Summarizing posterior inference: Binomial Example

- Output from Bayesian inference:

$$\text{posterior distribution } p(\theta|y) = \text{Beta}(y+1, n-y+1).$$

- Numerical summaries:
  1. mean $= \frac{y+1}{n+2}$
  2. median $\approx \frac{y+2/3}{n+4/3}$
  3. mode $= \frac{y}{n}$
  4. standard deviation

Summarizing posterior inference: Binomial Example

- Output from Bayesian inference:

$$\text{posterior distribution } p(\theta|y) = \text{Beta}(y+1, n-y+1).$$

- Numerical summaries:
  1. mean $= \frac{y+1}{n+2}$
  2. median $\approx \frac{y+2/3}{n+4/3}$
  3. mode $= \frac{y}{n}$
  4. standard deviation $= \sqrt{\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}}$
  5. interquartile range

Summarizing posterior inference: Binomial Example

- Output from Bayesian inference:

$$\text{posterior distribution } p(\theta|y) = \text{Beta}(y+1, n-y+1).$$

- Numerical summaries:
  1. mean $= \frac{y+1}{n+2}$
  2. median $\approx \frac{y+2/3}{n+4/3}$
  3. mode $= \frac{y}{n}$
  4. standard deviation $= \sqrt{\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}}$
  5. interquartile range

- Posterior interval $(100(1-\alpha)\%)$:

Summarizing posterior inference: Binomial Example

- Output from Bayesian inference:

$$\text{posterior distribution } p(\theta|y) = \text{Beta}(y+1, n-y+1).$$

- Numerical summaries:
  1. mean $= \frac{y+1}{n+2}$
  2. median $\approx \frac{y+2/3}{n+4/3}$
  3. mode $= \frac{y}{n}$
  4. standard deviation $= \sqrt{\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}}$
  5. interquartile range

- Posterior interval $(100(1-\alpha)\%)$: $[a, b]$, s.t. $\int_a^b \frac{\theta^y(1-\theta)^{n-y}}{B(y+1,n-y+1)} = 1-\alpha$

Freq vs. Bayes: Binomial Example

	Frequentist Inference	Bayesian Inference
Estimator	$\hat{\theta} = \frac{y}{n}$ (MLE)	$\frac{y+1}{n+2}$ (Posterior mean)
Variability	$\frac{\hat{\theta}(1-\hat{\theta})}{n}$ (Asymptotically)	$\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}$ (Posterior variance)
Interval	$[\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}]$ $\approx$ Confidence Interval	[a, b] Posterior Interval

Remark: $[a, b]$, s.t. $\int_a^b \frac{\theta^y(1-\theta)^{n-y}}{B(y+1,n-y+1)} = 0.95$.

C.I.: If confidence intervals are constructed using a given confidence level in an infinite number of independent experiments, the proportion of those intervals that contain the true value of the parameter will match the confidence level.

Informative prior distribution: Binomial Example

- Conjugate prior: $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$.

  Corresponds to $\alpha - 1$ prior successes and $\beta - 1$ prior failures.

Informative prior distribution: Binomial Example

- Conjugate prior: $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$.

  Corresponds to $\alpha - 1$ prior successes and $\beta - 1$ prior failures.

- Posterior: $p(\theta|y) \propto \text{Beta}(\theta|\alpha + y, \beta + n - y)$.

Informative prior distribution: Binomial Example

- Conjugate prior: $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$.

  Corresponds to $\alpha - 1$ prior successes and $\beta - 1$ prior failures.

- Posterior: $p(\theta|y) \propto \text{Beta}(\theta|\alpha + y, \beta + n - y)$.

- $E(\theta|y) = \frac{\alpha+y}{\alpha+\beta+n}$, $Var(\theta|y) = \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}$.

Informative prior distribution: Binomial Example

- Conjugate prior: $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$.

  Corresponds to $\alpha - 1$ prior successes and $\beta - 1$ prior failures.

- Posterior: $p(\theta|y) \propto \text{Beta}(\theta|\alpha + y, \beta + n - y)$.

- $E(\theta|y) = \frac{\alpha+y}{\alpha+\beta+n}$, $Var(\theta|y) = \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}$.

- What happens if $n \to \infty$?

Informative prior distribution: Binomial Example

- Conjugate prior: $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$.

  Corresponds to $\alpha - 1$ prior successes and $\beta - 1$ prior failures.

- Posterior: $p(\theta|y) \propto \text{Beta}(\theta|\alpha + y, \beta + n - y)$.

- $E(\theta|y) = \frac{\alpha+y}{\alpha+\beta+n}$, $Var(\theta|y) = \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}$.

- What happens if $n \to \infty$?

- Central limit theorem

$$\left(\frac{\theta - E(\theta|y)}{\sqrt{Var(\theta|y)}} \middle| y \right) \to N(0,1).$$

Plan

Normal distribution with known variance

- $y \sim N(\theta, \sigma^2)$, i.e. sampling distribution is

$$p(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\theta)^2}.$$

Normal distribution with known variance

- $y \sim N(\theta, \sigma^2)$, i.e. sampling distribution is

$$p(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\theta)^2}.$$

- Conjugate prior: $\theta \sim N(\mu_0, \tau_0^2)$, i.e.

$$p(\theta) \propto \exp\left(-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2\right).$$

Normal distribution with known variance

- $y \sim N(\theta, \sigma^2)$, i.e. sampling distribution is

$$p(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\theta)^2}.$$

- Conjugate prior: $\theta \sim N(\mu_0, \tau_0^2)$, i.e.

$$p(\theta) \propto \exp\left(-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2\right).$$

- Posterior distribution:

$$p(\theta|y) \propto \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2 - \frac{1}{2\tau_0^2}(\theta - \mu_0)^2\right).$$

Normal distribution with known variance

Posterior distribution

$$p(\theta|y) \propto \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2 - \frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right)$$
$$\propto \exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right),$$

where

$$\mu_1 = \frac{\mu_0/\tau_0^2 + y/\sigma^2}{1/\tau_0^2 + 1/\sigma^2}, \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}.$$

Normal distribution with known variance

Posterior distribution

$$p(\theta|y) \propto \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2 - \frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right)$$
$$\propto \exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right),$$

where

$$\mu_1 = \frac{\mu_0/\tau_0^2 + y/\sigma^2}{1/\tau_0^2 + 1/\sigma^2}, \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}.$$

Inverse variance — precision.

Normal distribution with known variance

Posterior distribution

$$p(\theta|y) \propto \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2 - \frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right)$$
$$\propto \exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right),$$

where

$$\mu_1 = \frac{\mu_0/\tau_0^2 + y/\sigma^2}{1/\tau_0^2 + 1/\sigma^2}, \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}.$$

Inverse variance — precision.

Posterior precision = prior precision + data precision.

Normal distribution with known variance

Posterior distribution: Mean

$$\mu_1 = \frac{\mu_0/\tau_0^2 + y/\sigma^2}{1/\tau_0^2 + 1/\sigma^2}.$$

Normal distribution with known variance

Posterior distribution: Mean

$$\mu_1 = \frac{\mu_0/\tau_0^2 + y/\sigma^2}{1/\tau_0^2 + 1/\sigma^2}.$$

Shrinkage

Normal distribution with known variance

Posterior distribution: Mean

$$\mu_1 = \frac{\mu_0/\tau_0^2 + y/\sigma^2}{1/\tau_0^2 + 1/\sigma^2}.$$

Shrinkage

- prior mean 'shrinks' toward observation

$$\mu_1 = \mu_0 + (y - \mu_0)\frac{\tau_0^2}{\sigma^2 + \tau_0^2}.$$

Normal distribution with known variance

Posterior distribution: Mean

$$\mu_1 = \frac{\mu_0/\tau_0^2 + y/\sigma^2}{1/\tau_0^2 + 1/\sigma^2}.$$

Shrinkage

- prior mean 'shrinks' toward observation

$$\mu_1 = \mu_0 + (y - \mu_0)\frac{\tau_0^2}{\sigma^2 + \tau_0^2}.$$

- data 'shrinks' toward prior mean

$$\mu_1 = y - (y - \mu_0)\frac{\sigma^2}{\sigma^2 + \tau_0^2}.$$

Normal distribution with known variance

Normal distribution with known variance

Posterior distribution: Variance

$$\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}.$$

Normal distribution with known variance

Posterior distribution: Variance

$$\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}.$$

Inverse variance — precision.

Normal distribution with known variance

Posterior distribution: Variance

$$\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}.$$

Inverse variance — precision.

Posterior precision = prior precision + data precision.

Normal distribution with known variance

Posterior Predictive Distribution

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

Normal distribution with known variance

Posterior Predictive Distribution

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

Recall: $E(\tilde{y}|\theta) = \theta$, $Var(\tilde{y}|\theta) = \sigma^2$. Then

Normal distribution with known variance

Posterior Predictive Distribution

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

Recall: $E(\tilde{y}|\theta) = \theta$, $Var(\tilde{y}|\theta) = \sigma^2$. Then

$$E(\tilde{y}|y) = E[E(\tilde{y}|\theta, y)|y] = E(\theta|y) = \mu_1;$$

$$\begin{aligned} Var(\tilde{y}|y) &= E(Var(\tilde{y}|\theta, y)|y) + Var(E(\tilde{y}|\theta, y)|y) \\ &= E(\sigma^2|y) + Var(\theta|y) = \sigma^2 + \tau_1^2. \end{aligned}$$

Normal distribution with known variance

Normal model with multiple observations

$$p(\theta|y) \propto p(\theta)p(y|\theta) = p(\theta)\prod_{i=1}^{n} p(y|\theta)$$

$$\propto \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right)\prod_{i=1}^{n}\exp\left(-\frac{(y_i-\theta)^2}{2\sigma^2}\right)\ldots$$

Normal distribution with known variance

Normal model with multiple observations

$$p(\theta|y) \propto p(\theta)p(y|\theta) = p(\theta)\prod_{i=1}^{n} p(y|\theta)$$

$$\propto \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right)\prod_{i=1}^{n}\exp\left(-\frac{(y_i-\theta)^2}{2\sigma^2}\right)\ldots$$

Let $\bar{y} = \sum_{i=1}^{n} y_i/n$, then $p(\theta|y) = N(\theta|\mu_n, \tau_n^2)$,

$$\mu_n = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}, \quad \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}.$$

Plan

Other Examples

- Normal distribution with known mean but unknown variance

$$y_i \overset{i.i.d.}{\sim} N(0, \sigma^2), 1 \leq i \leq n; \quad p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} \exp(-\beta/\sigma^2).$$

Other Examples

- Normal distribution with known mean but unknown variance

$$y_i \overset{i.i.d.}{\sim} N(0, \sigma^2), 1 \leq i \leq n; \quad p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} \exp(-\beta/\sigma^2).$$

- Poisson model with Gamma prior

$$y_i \overset{i.i.d.}{\sim} \text{Pois}(\theta), 1 \leq i \leq n; \quad p(\theta) \propto \exp(-\beta\theta)\theta^{\alpha-1}.$$

Other Examples

- Normal distribution with known mean but unknown variance

$$y_i \overset{i.i.d.}{\sim} N(0, \sigma^2), 1 \leq i \leq n; \quad p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} \exp(-\beta/\sigma^2).$$

- Poisson model with Gamma prior

$$y_i \overset{i.i.d.}{\sim} \text{Pois}(\theta), 1 \leq i \leq n; \quad p(\theta) \propto \exp(-\beta\theta)\theta^{\alpha-1}.$$

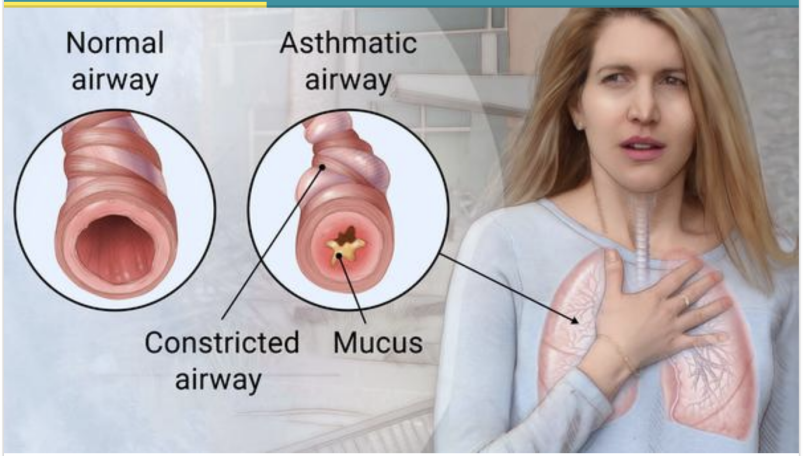
- Exponential model with Gamma prior.

Example: Asthma

Asthma

Also called: bronchial asthma

ABOUT SYMPTOMS TREATMENTS



A condition in which a person's airways become inflamed, narrow and swell, and produce extra mucus, which makes it difficult to breathe.

Very common

More than 3 million US cases per year

Asthma can be minor or it can interfere with daily activities. In some cases, it may lead to a life-threatening attack.

Asthma may cause difficulty breathing, chest pain, cough, and wheezing. The symptoms may sometimes flare-up.

Asthma can usually be managed with rescue inhalers to treat symptoms (albuterol) and controller inhalers that prevent symptoms (steroids). Severe cases may require longer-acting inhalers that keep the airways open (formoterol, salmeterol, tiotropium), as well as oral steroids.

Ages affected

Example: Asthma

- Observe: 3 out of $200,000$ in a city in US.

Example: Asthma

- Observe: 3 out of $200,000$ in a city in US.
- Poisson sampling model

$$y \sim \text{Poisson}(x\theta), x = 2.$$

θ represents true underlying long-term asthma mortality rate.

Example: Asthma

- Observe: 3 out of $200,000$ in a city in US.
- Poisson sampling model

$$y \sim \text{Poisson}(x\theta), x = 2.$$

  θ represents true underlying long-term asthma mortality rate.

- Setting up a prior distribution: ≈ 0.6, ≤ 1.5 per 100,00 people.

Example: Asthma

- Observe: 3 out of $200,000$ in a city in US.
- Poisson sampling model

$$y \sim \text{Poisson}(x\theta), x = 2.$$

θ represents true underlying long-term asthma mortality rate.

- Setting up a prior distribution: ≈ 0.6, ≤ 1.5 per 100,00 people.

$$\theta \sim \text{Gamma}(3.0, 5.0).$$

- Observation: $y = 3, x = 2$.

Example: Asthma

- Prior

$$\theta \sim \text{Gamma}(3.0, 5.0).$$

- Observation: $y = 3, x = 2$.

Example: Asthma

- Prior

$$\theta \sim \text{Gamma}(3.0, 5.0).$$

- Observation: $y = 3, x = 2$.

- Posterior

$$\theta \sim \text{Gamma}(6.0, 7.0).$$

Example: Asthma

- Prior

$$\theta \sim \text{Gamma}(3.0, 5.0).$$

- Observation: $y = 3, x = 2$.
- Posterior

$$\theta \sim \text{Gamma}(6.0, 7.0).$$

- What if we observe 30 deaths over 10 years?

Example: cancer rates

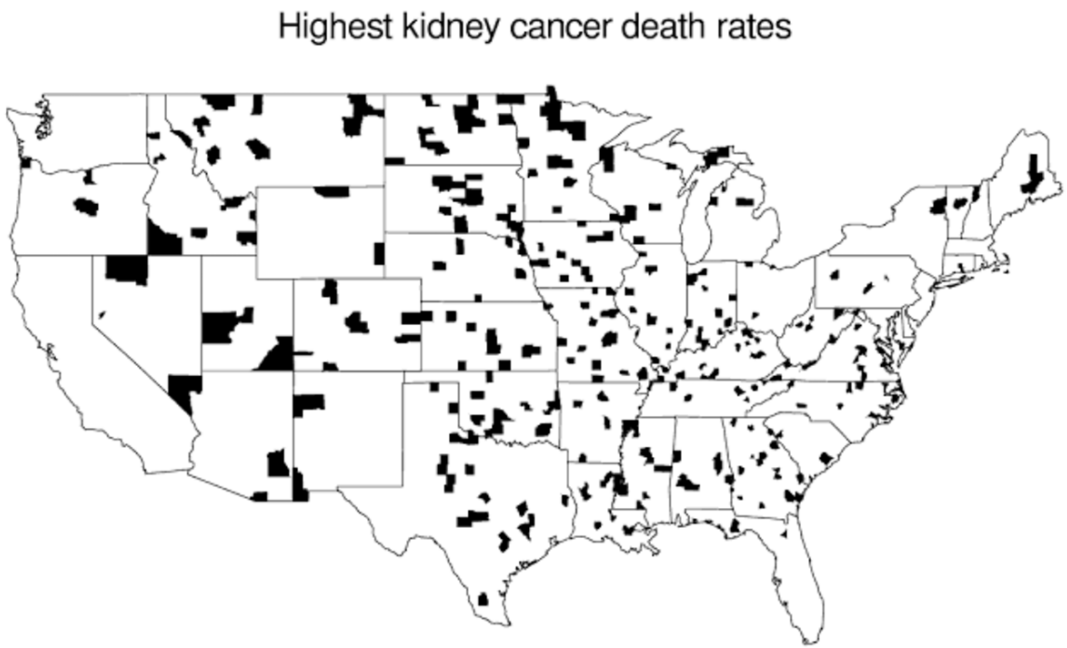


Figure 2.6 *The counties of the United States with the highest 10% age-standardized death rates for cancer of kidney/ureter for U.S. white males, 1980–1989. Why are most of the shaded counties in the middle of the country? See Section 2.7 for discussion.*

Example: cancer rates

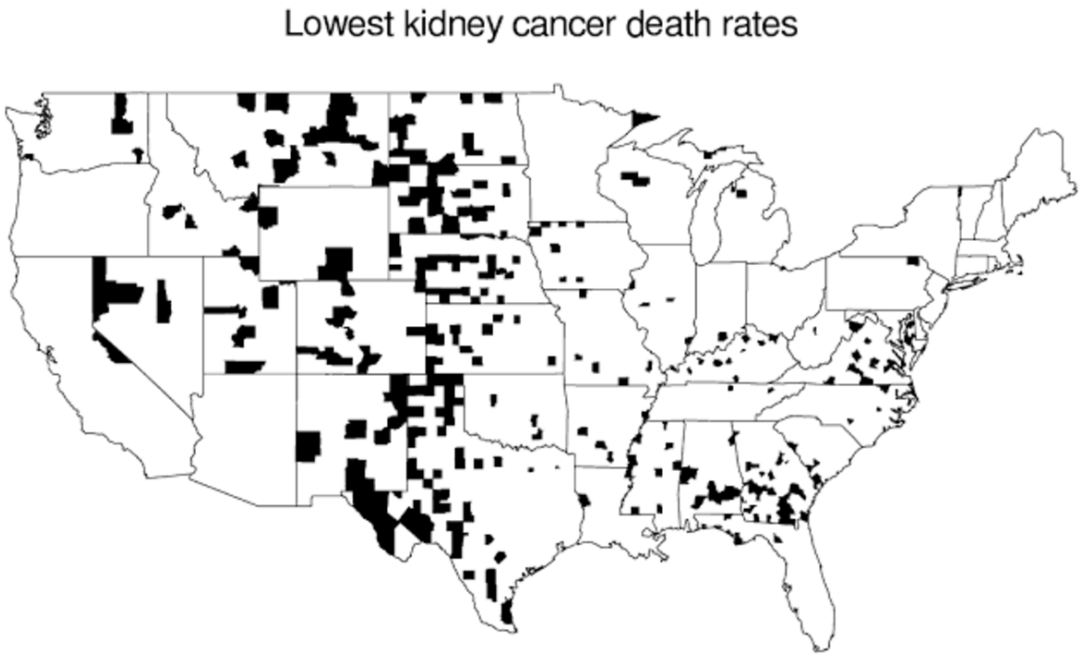


Figure 2.7 *The counties of the United States with the lowest 10% age-standardized death rates for cancer of kidney/ureter for U.S. white males, 1980–1989. Surprisingly, the pattern is somewhat similar to the map of the highest rates, shown in Figure 2.6.*

Example: cancer rates

- Some counties in Great Plains have the lowest & highest rates.

Example: cancer rates

- Some counties in Great Plains have the lowest & highest rates.
- Why?

Example: cancer rates

- Some counties in Great Plains have the lowest & highest rates.
- Why?
- Sample Size.

Example: cancer rates

- Some counties in Great Plains have the lowest & highest rates.
- Why?
- Sample Size.
  - Kidney cancer is a rare disease.

Example: cancer rates

- Some counties in Great Plains have the lowest & highest rates.
- Why?
- Sample Size.
  - Kidney cancer is a rare disease.
  - Example: county of population 1000.

Example: cancer rates

- Some counties in Great Plains have the lowest & highest rates.
- Why?
- Sample Size.
  - Kidney cancer is a rare disease.
  - Example: county of population 1000.
  - Great Plains have many low-population counties.

Example: cancer rates

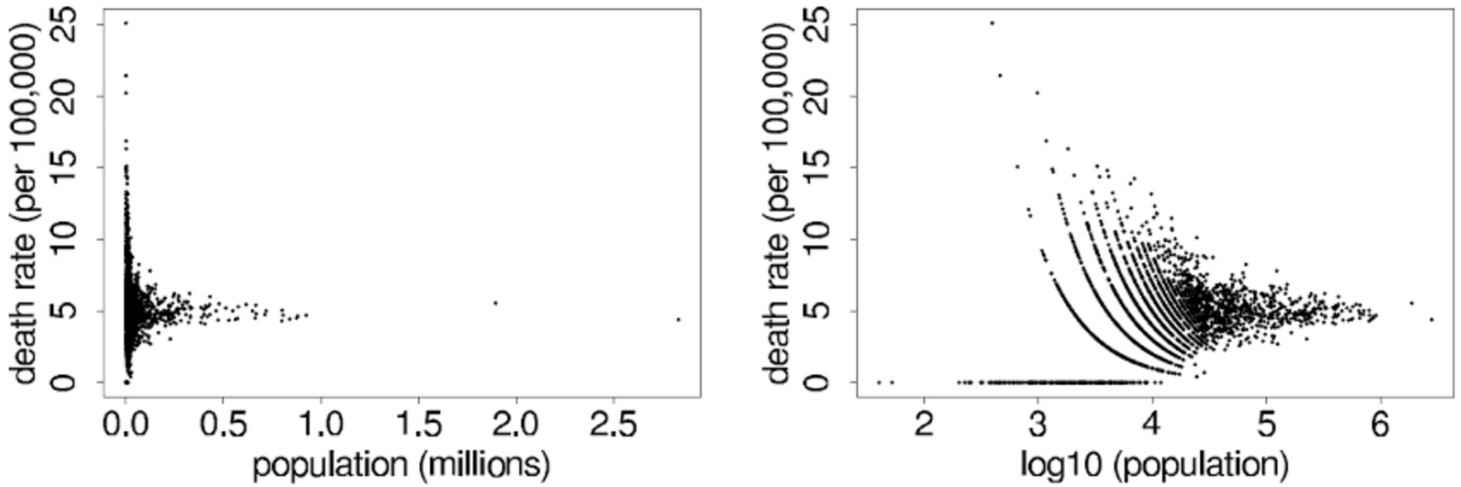


Figure 2.8 *(a) Kidney cancer death rates $y_j/(10n_j)$ vs. population size n_j. (b) Replotted on the scale of* $\log_{10}$ *population to see the data more clearly. The patterns come from the discreteness of the data* ($n_j = 0, 1, 2, \ldots$).

Example: cancer rates

Bayesian model

- Notation:
  - y_j: number of kidney cancer death in county j from 1980-1989,
  - θ_j: underlying rate in units of deaths per person per year.
- Likelihood:

$$y_j \sim \text{Poisson}(10 n_j \theta_j).$$

Example: cancer rates

Bayesian model

- Notation:
  - y_j: number of kidney cancer death in county j from 1980-1989,
  - θ_j: underlying rate in units of deaths per person per year.
- Likelihood:

$$y_j \sim \text{Poisson}(10 n_j \theta_j).$$

- Informative prior:
  - Gamma distribution with $\alpha = 20$ and $\beta = 430,000$.
  - Prior mean: $\frac{\alpha}{\beta} = 4.65 \times 10^{-5}$, standard deviation: $\frac{\sqrt{\alpha}}{\beta} = 1.04 \times 10^{-5}$.

Example: cancer rates

Bayesian model

- Notation:
  - y_j: number of kidney cancer death in county j from 1980-1989,
  - θ_j: underlying rate in units of deaths per person per year.
- Likelihood:

$$y_j \sim \text{Poisson}(10 n_j \theta_j).$$

- Informative prior:
  - Gamma distribution with $\alpha = 20$ and $\beta = 430,000$.
  - Prior mean: $\frac{\alpha}{\beta} = 4.65 \times 10^{-5}$, standard deviation: $\frac{\sqrt{\alpha}}{\beta} = 1.04 \times 10^{-5}$.
  - Posterior:

$$\theta_j | y_j \sim \text{Gamma}(20 + y_j, 430,000 + 10 n_j).$$

Example: cancer rates

Inference for a small county

- $n_j = 1000$
- raw death rate v.s. posterior mean ($y_j = 0, 1, 2$)
- How likely, a priori, is that y_j will equal $0, 1, 2$?

Example: cancer rates

Inference for a small county

- $n_j = 1000$
- raw death rate v.s. posterior mean ($y_j = 0, 1, 2$)
- How likely, a priori, is that y_j will equal $0, 1, 2$?
  1. Draw 500 θ_j from prior
  2. For each θ_j, draw y_j from Poisson with $n_j\theta_j$.

Example: cancer rates

Inference for a small county

- $n_j = 1000$
- raw death rate v.s. posterior mean ($y_j = 0, 1, 2$)
- How likely, a priori, is that y_j will equal $0, 1, 2$?
  1. Draw 500 θ_j from prior
  2. For each θ_j, draw y_j from Poisson with $n_j\theta_j$.
  3. Out of 500 y_j, 319 are 0, 141 are 1 and 33 are 2.

Example: cancer rates

Inference for a small county

- $n_j = 1000$
- raw death rate v.s. posterior mean ($y_j = 0, 1, 2$)
- How likely, a priori, is that y_j will equal $0, 1, 2$?
  1. Draw 500 θ_j from prior
  2. For each θ_j, draw y_j from Poisson with $n_j\theta_j$.
  3. Out of 500 y_j, 319 are 0, 141 are 1 and 33 are 2.

Inference for a big county

- $n_j = 1$ million
- How many cancer deaths y_j might we expect in a 10-year period?

Example: cancer rates

Inference for a small county

- $n_j = 1000$
- raw death rate v.s. posterior mean ($y_j = 0, 1, 2$)
- How likely, a priori, is that y_j will equal $0, 1, 2$?
  1. Draw 500 θ_j from prior
  2. For each θ_j, draw y_j from Poisson with $n_j\theta_j$.
  3. Out of 500 y_j, 319 are 0, 141 are 1 and 33 are 2.

Inference for a big county

- $n_j = 1$ million
- How many cancer deaths y_j might we expect in a 10-year period?
  - Raw death rate (simulation): 50% interval $[3.93 \times 10^{-5}, 5.45 \times 10^{-5}]$.

Example: cancer rates

Inference for a small county

- $n_j = 1000$
- raw death rate v.s. posterior mean ($y_j = 0, 1, 2$)
- How likely, a priori, is that y_j will equal $0, 1, 2$?
  1. Draw 500 θ_j from prior
  2. For each θ_j, draw y_j from Poisson with $n_j\theta_j$.
  3. Out of 500 y_j, 319 are 0, 141 are 1 and 33 are 2.

Inference for a big county

- $n_j = 1$ million
- How many cancer deaths y_j might we expect in a 10-year period?
  - Raw death rate (simulation): 50% interval $[3.93 \times 10^{-5}, 5.45 \times 10^{-5}]$.
  - Corresponding Bayesian estimate: 50% interval $[3.96 \times 10^{-5}, 5.41 \times 10^{-5}]$.

Example: cancer rates

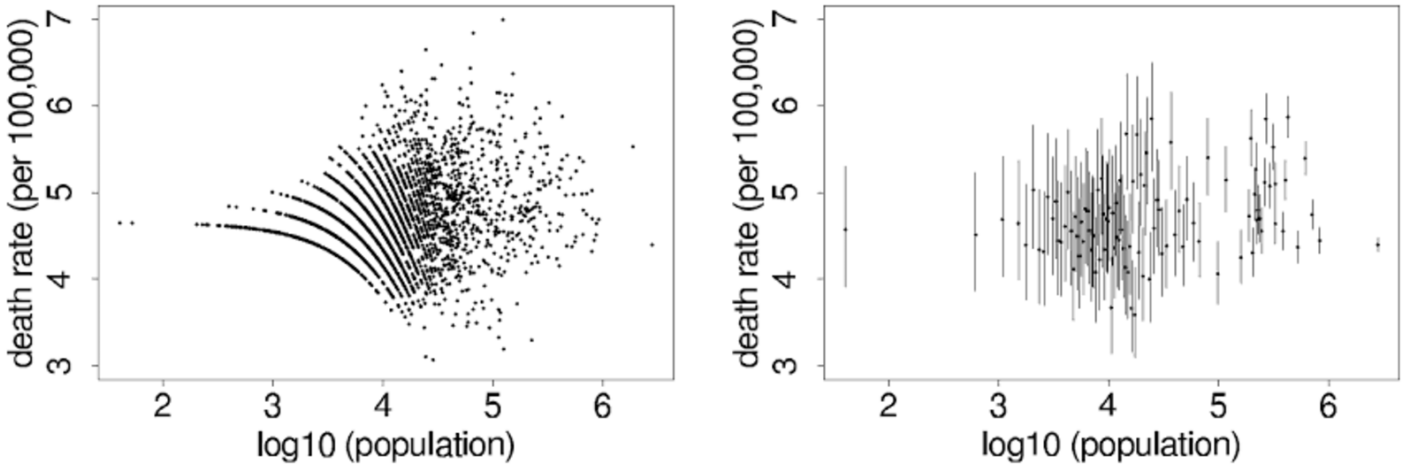


Figure 2.9 *(a) Bayes-estimated posterior mean kidney cancer death rates,* $E(\theta_j|y_j) = \frac{20+y_j}{430{,}000+10n_j}$ *vs. logarithm of population size* n_j*, the 3071 counties in the U.S. (b) Posterior medians and 50% intervals for* θ_j *for a sample of 100 counties* j*. The scales on the y-axes differ from the plots in Figure 2.8b.*

Example: cancer rates

How prior is constructed?

Example: cancer rates

How prior is constructed?

- Recall: $\theta_j \sim \text{Gamma}(20, 430000)$.

Example: cancer rates

How prior is constructed?

- Recall: $\theta_j \sim$ Gamma$(20, 430000)$.
- Predictive distribution

$$p(y_j) = \int p(y_j|\theta_j)p(\theta_j)d\theta_j = \text{Neg-bin}(\alpha, \frac{\beta}{10n_j}).$$

Example: cancer rates

How prior is constructed?

- Recall: $\theta_j \sim \text{Gamma}(20, 430000)$.
- Predictive distribution

$$p(y_j) = \int p(y_j|\theta_j)p(\theta_j)d\theta_j = \text{Neg-bin}(\alpha, \frac{\beta}{10n_j}).$$

- Determine α, β through

$$E\left(\frac{y_j}{10n_j}\right) = \frac{\alpha}{\beta}, Var\left(\frac{y_j}{10n_j}\right) = \frac{1}{10n_j}\frac{\alpha}{\beta} + \frac{\alpha}{\beta^2}.$$

Substitute the expectations/variances with sample expectations/variances.

Example: cancer rates

How prior is constructed?

- Recall: $\theta_j \sim \text{Gamma}(20, 430000)$.
- Predictive distribution

$$p(y_j) = \int p(y_j|\theta_j)p(\theta_j)d\theta_j = \text{Neg-bin}(\alpha, \frac{\beta}{10n_j}).$$

- Determine α, β through

$$E\left(\frac{y_j}{10n_j}\right) = \frac{\alpha}{\beta}, Var\left(\frac{y_j}{10n_j}\right) = \frac{1}{10n_j}\frac{\alpha}{\beta} + \frac{\alpha}{\beta^2}.$$

  Substitute the expectations/variances with sample expectations/variances.
- Later in the course: more principled way through hierarchical model.

Example: cancer rates

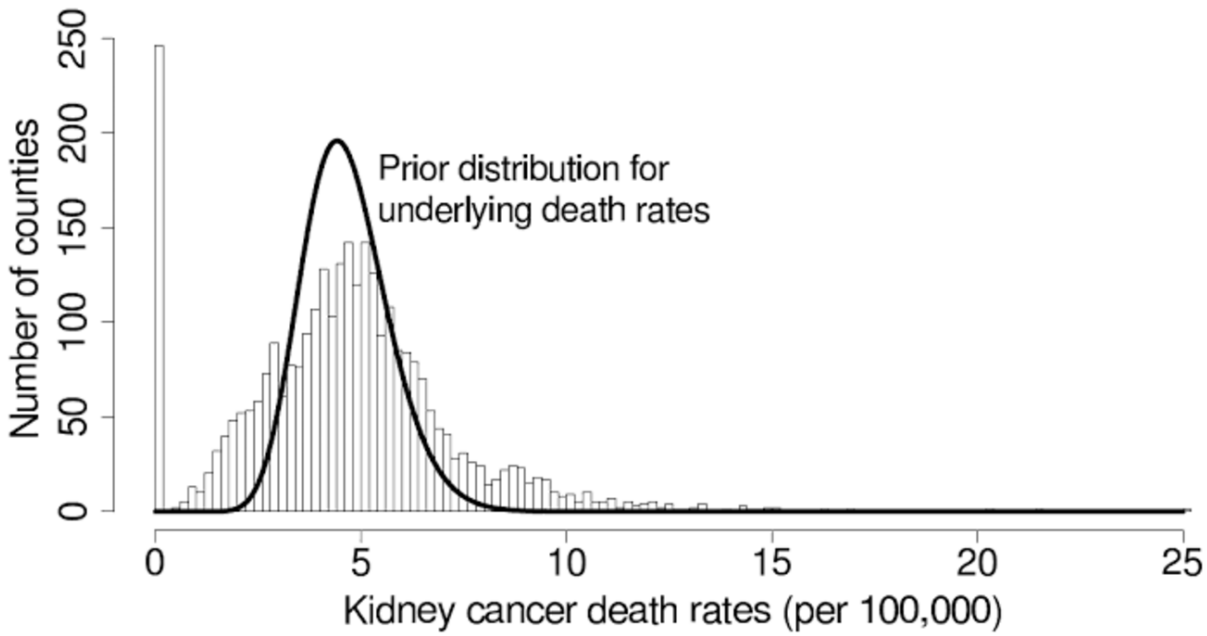
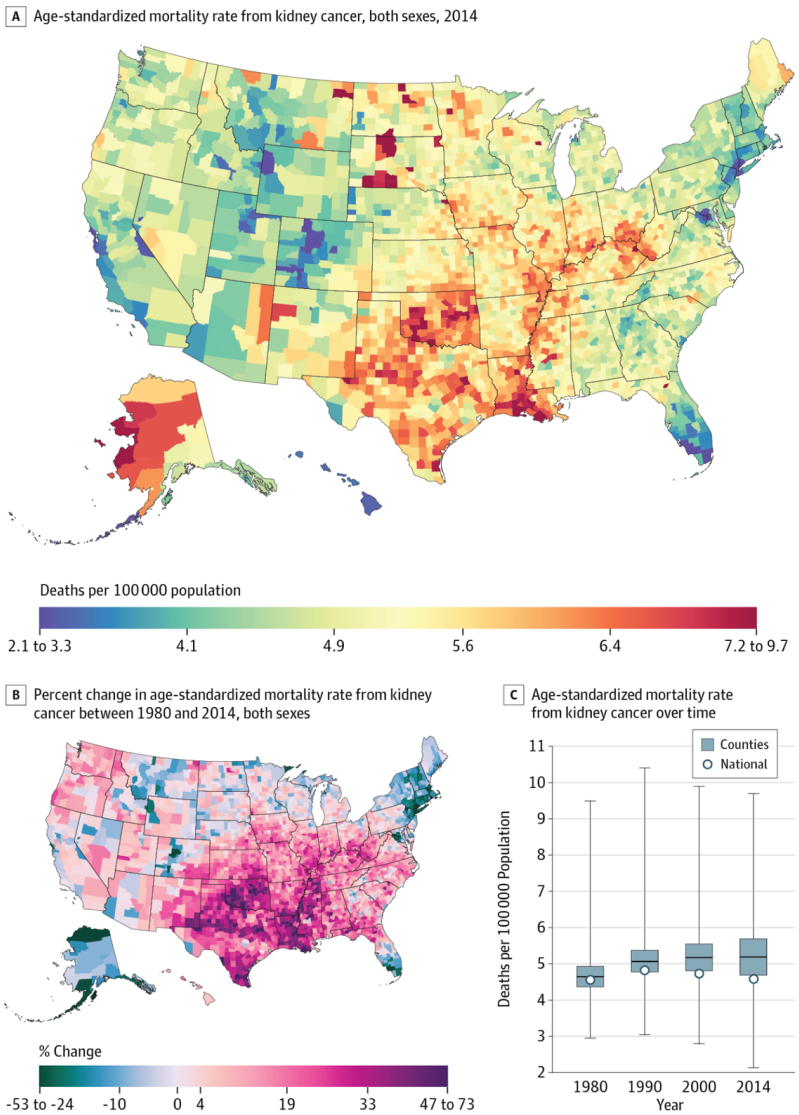


Figure 2.10 *Empirical distribution of the age-adjusted kidney cancer death rates,* $\frac{y_j}{10n_j}$*, for the 3071 counties in the U.S., along with the Gamma*$(20, 430{,}000)$ *prior distribution for the underlying cancer rates* θ_j*.*

Further reading

Cancer Mortality

Mokdad AH, et al. Trends and Patterns of Disparities in Cancer Mortality Among US Counties, 1980-2014. JAMA. 2017;317(4):388406.



Plan

Discussions on priors

- Noninformative prior distributions
  'Let the data speak for themselves.'

Discussions on priors

- Noninformative prior distributions
  'Let the data speak for themselves.'
- Weakly informative prior distributions

Discussions on priors

- Noninformative prior distributions
  'Let the data speak for themselves.'
- Weakly informative prior distributions
- Proper and improper prior distributions

Discussions on priors

- Noninformative prior distributions
  'Let the data speak for themselves.'
- Weakly informative prior distributions
- Proper and improper prior distributions
  - Is the corresponding posterior proper?

Discussions on priors

- Noninformative prior distributions
  'Let the data speak for themselves.'
- Weakly informative prior distributions
- Proper and improper prior distributions
  - Is the corresponding posterior proper?
  - Example: Gaussian model with $p(\sigma^2) \propto 1/\sigma^2$.

Discussions on priors

- Noninformative prior distributions
  'Let the data speak for themselves.'
- Weakly informative prior distributions
- Proper and improper prior distributions
  - Is the corresponding posterior proper?
  - Example: Gaussian model with $p(\sigma^2) \propto 1/\sigma^2$.
- Jeffreys' prior: $p(\theta) \propto |J(\theta)|^{1/2}$, $J(\theta)$ is Fisher information.

Discussions on priors

- Noninformative prior distributions
  'Let the data speak for themselves.'
- Weakly informative prior distributions
- Proper and improper prior distributions
  - Is the corresponding posterior proper?
  - Example: Gaussian model with $p(\sigma^2) \propto 1/\sigma^2$.
- Jeffreys' prior: $p(\theta) \propto |J(\theta)|^{1/2}$, $J(\theta)$ is Fisher information.
  - Jeffreys' invariance principle: $p(\theta)p(y|\theta) \Leftrightarrow p(\phi)p(y|\phi)$.

Discussions on priors

- Noninformative prior distributions
  'Let the data speak for themselves.'
- Weakly informative prior distributions
- Proper and improper prior distributions
  - Is the corresponding posterior proper?
  - Example: Gaussian model with $p(\sigma^2) \propto 1/\sigma^2$.
- Jeffreys' prior: $p(\theta) \propto |J(\theta)|^{1/2}$, $J(\theta)$ is Fisher information.
  - Jeffreys' invariance principle: $p(\theta)p(y|\theta) \Leftrightarrow p(\phi)p(y|\phi)$.
  - Invariant to parameterization.

$$J(\phi) = J(\theta)|d\theta/d\phi|^2.$$

Discussions on priors: Example

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

Discussions on priors: Example

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

- Jeffreys' prior
$$p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2},$$
which is Beta$(1/2, 1/2)$.

Discussions on priors: Example

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

- Jeffreys' prior
$$p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2},$$
which is Beta$(1/2, 1/2)$.

- Flat prior $p(\theta) \propto 1$, which is Beta$(1, 1)$.

Discussions on priors: Example

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

- Jeffreys' prior
$$p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2},$$
which is Beta$(1/2, 1/2)$.
- Flat prior $p(\theta) \propto 1$, which is Beta$(1, 1)$.
- $p(\textit{logit}(\theta)) \propto 1$, which is Beta$(0, 0)$.

Discussions on priors: Example

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

- Jeffreys' prior
$$p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2},$$
which is Beta$(1/2, 1/2)$.
- Flat prior $p(\theta) \propto 1$, which is Beta$(1, 1)$.
- $p(\textit{logit}(\theta)) \propto 1$, which is Beta$(0, 0)$.
  - Improper when $y = 0$ or n.