Parameter Estimation
Fitting Probability Distributions
Bayesian Approach

MIT 18.443

Dr. Kempthorne

Spring 2015

Outline

1. Bayesian Approach to Parameter Estimation
   - Framework/Definitions/Examples

Bayesian Framework: Extension of Maximum Likelihood

General Model

- Data Model : $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ vector-valued random variable with joint density given by

  $$f(x_1, \ldots, x_n \mid \theta)$$

- Data Realization: $\mathbf{X} = \mathbf{x} = (x_1, \ldots, x_n)$
- **Likelihood of** θ (given $\mathbf{x}$):

  $$lik(\theta) = f(x_1, \ldots, x_n \mid \theta)$$

  (MLE $\hat{\theta}$ maximizes $lik(\theta)$ for fixed realization)
- **Prior distribution:** true $\theta \in \Theta$ modeled as random variable

  $$\theta \sim \Pi, \text{ with density } \pi(\theta), \theta \in \Theta$$

- **Posterior Distribution:** Distribution of θ given $\mathbf{X} = \mathbf{x}$
  - Joint density of (X, θ): $f_{\mathbf{X},\theta}(\mathbf{x}, \theta) = f(\mathbf{x} \mid \theta)\pi(\theta)$
  - Density of marginal distribution of $\mathbf{X}$:

    $$f_{\mathbf{X}}(\mathbf{x}) = \int_\Theta f_{\mathbf{X},\theta}(\mathbf{x}, \theta) d\theta = \int_\Theta f(\mathbf{x} \mid \theta)\pi(\theta) d\theta$$

  - Density of posterior distribution of θ given $\mathbf{X} = \mathbf{x}$

    $$\pi(\theta \mid \mathbf{x}) = \frac{f_{\mathbf{X},\theta}(\mathbf{x}, \theta)}{f_{\mathbf{X}}(\mathbf{x})}$$

Bayesian Framework

Posterior Distribution: Conditional distribution of θ given $\mathbf{X} = \mathbf{x}$

$$\pi(\theta \mid \mathbf{x}) = \frac{f_{\mathbf{X},\theta}(\mathbf{x},\theta)}{f_{\mathbf{X}}(\mathbf{x})} = \frac{f(\mathbf{x} \mid \theta)\pi(\theta)}{\int_\Theta f(\mathbf{x} \mid \theta)\pi(\theta)d\theta}$$

$$\propto \quad = f(\mathbf{x} \mid \theta)\pi(\theta)$$

$$\text{Posterior density} \quad \propto \quad = \mathit{Likelihood}(\theta) \times \text{ Prior density}$$

Bayesian Principles

- Prior distribution models uncertainty about θ, a priori (before observing any data)
- Justified by axioms of statistical decision theory (utility theory and the optimality of maximizing expected utility).
- All information about θ is contained in $\pi(\theta \mid \mathbf{x})$
- Posterior mean minimizes expected squared error
  $$E[(\theta - a)^2 \mid \mathbf{x}] \text{ minimized by } a = E[\theta \mid \mathbf{x}].$$
- Posterior median minimizes expected absolute error
  $$E[|\theta - a| \mid \mathbf{x}] \text{ minimized by } a = \mathit{median}(\theta \mid \mathbf{x}).$$

Bayesian Framework

Bayesian Principles (continued):

- **Posterior Mode**: Modal value of $\pi(\theta \mid \mathbf{x})$ is most probable.
- Analogue to 90% confidence interval: θ values between 0.05 and 0.95 quantiles of $\pi(\theta \mid \mathbf{x})$.
- **Highest posterior density (HPD) interval (region)**:

  For $\alpha : 0 < \alpha < 1$, the $(1-\alpha)HPD$ region for θ is
  $$R_{d^*} = \{\theta : \pi(\theta \mid \mathbf{x}) > d^*\}$$
  where d^* is the value such that $\pi(R_{d^*} \mid \mathbf{x}) = 1 - \alpha$.

  Note: if posterior density is unimodal but not symmetric, then the tail probabilities outside the region will be unequal.

Bayesian Inference: Bernoulli Trials

Bernoulli Trials: $X_1, X_2, \ldots, X_n$ i.i.d. *Bernoulli*(θ)

- Sample Space: $\mathcal{X} = \{1, 0\}$ ("success" or "failure")
- Probability mass function

$$f(x \mid \theta) = \begin{cases} \theta & , \textit{if} \quad x = 1 \\ (1-\theta) & , \textit{if} \quad x = 0 \end{cases}$$

Examples:

- Flipping a coin and observing a *Head* versus a *Tail*.
- Random sample from a population and measuring a dichotomous attribute (e.g., preference for a given political candidate, testing positive for a given disease) .

Summary Statistic: $S = X_1 + X_2 + \cdots + X_n$

$$S \sim \textit{Binomial}(n, p)$$

$$P(S = k \mid \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}, \; k = 0, 1, \ldots, n.$$

Bayesian Inference: Bernoulli Trials

Case 1: Uniform Prior for $\theta \in \Theta = \{\theta : 0 \leq \theta \leq 1\} = [0, 1]$

- Prior density for θ:
$$\pi(\theta) = 1,\ 0 \leq \theta \leq 1$$
- Joint density/pmf for (S, θ)
$$\begin{aligned} f_{S,\theta}(s,\theta) &= f_{S|\theta}(s \mid \theta)\pi(\theta) \\ &= \binom{n}{s} \theta^s (1-\theta)^{(n-s)} \times 1 \end{aligned}$$
- Marginal density of S
$$\begin{aligned} f_S(s) &= \int_0^1 \binom{n}{s} \theta^s (1-\theta)^{(n-s)} d\theta \\ &= \binom{n}{s} \int_0^1 \theta^s (1-\theta)^{(n-s)} d\theta \\ &= \binom{n}{s} Beta(s+1, (n-s)+1) \equiv \tfrac{1}{n+1} \end{aligned}$$
- Posterior density of θ given S
$$\pi(\theta \mid s) = f_{S,\theta}(s,\theta)/f_S(s)$$

Bayesian Inference: Bernoulli Trials

Case 1: Uniform Prior (continued)

- Posterior density of θ given S

$$\begin{aligned} \pi(\theta \mid s) &= f_{S,\theta}(s,\theta)/f_S(s) \\ &= \frac{\theta^s(1-\theta)^{(n-s)}}{Beta(s+1,(n-s)+1)} \end{aligned}$$

Recall a random variable $U \sim Beta(a,b)$, has density

$$g(u \mid a,b) = \frac{u^{a-1}(1-u)^{b-1}}{Beta(a,b)}, 0 < u < 1$$

where

$$Beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \text{ with}$$

$$\Gamma(a) = \int_0^\infty y^{\alpha-1}e^{-x}dx, \text{ (see } Gamma(a) \text{ density)}$$

$$\Gamma(a+1) = a \times \Gamma(a) = (a!) \text{ for integral } a$$

Also (Appendix A3 of Rice, 2007)

$$\begin{aligned} E[U \mid a,b] &= a/(a+b) \\ Var[U \mid a,b] &= ab/[(a+b)^2(a+b+1)] \end{aligned}$$

Bayesian Inference: Bernoulli Trials

Case 1: Uniform Prior (continued)

- Prior: $\theta \sim Beta(a = 1, b = 1)$, a priori
- Sample data: $n = 20$ and $S = \sum_{i=1}^{n} X_i = 13$ (Example 3.5.E)
- Posterior: $[\theta \mid S = s] \sim Beta(a, b)$ with

$$a = s + 1 = 14 \text{ and } b = (n - s) + 1 = 8$$

  Use R to compute:

  - Posterior mean: $a/(a + b)$
  - Posterior standard deviation: $\sqrt{ab/[(a + b)^2(a + b + 1)]}$
  - Posterior probability:$\pi(\{\theta \leq .5\} \mid s)$

```
> a=14; b=8
> a/(a+b)
[1] 0.6363636
> sqrt(a*b/(((a+b)**2)*(a+b +1)))
[1] 0.100305
> pbeta(.5,shape1=14, shape2=8)
[1] 0.09462357
```

Bayesian Inference: Bernoulli Trials

Case 2: Beta Prior for $\theta \in \Theta = \{\theta : 0 \leq \theta \leq 1\} = [0, 1]$

- Prior density for θ:
$$\pi(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{Beta(a,b)}, \ 0 \leq \theta \leq 1$$
- Joint density/pmf for (S, θ)
$$\begin{aligned} f_{S,\theta}(s, \theta) &= f_{S|\theta}(s \mid \theta)\pi(\theta) \\ &= \binom{n}{s} \theta^s (1-\theta)^{(n-s)} \times \frac{\theta^{a-1}(1-\theta)^{b-1}}{Beta(a,b)} \\ &\propto \theta^{s+a-1}(1-\theta)^{(n-s)+b-1} \end{aligned}$$
- Posterior density of θ given S
$$\begin{aligned} \pi(\theta \mid s) &= f_{S,\theta}(s, \theta)/f_S(s) \\ &= \frac{\theta^{s+a-1}(1-\theta)^{(n-s)+b-1}}{\int_{\theta'} (\theta')^{s+a-1}(1-\theta')^{(n-s)+b-1} d\theta'} \\ &= \frac{\theta^{s+a-1}(1-\theta)^{(n-s)+b-1}}{Beta((s+a-1, (n-s)+b-1)} \end{aligned}$$

Bayesian Inference: Bernoulli Trials

Case 2: Beta Prior (continued)

- Posterior density of θ given S

$$\begin{aligned}\pi(\theta \mid s) &= f_{S,\theta}(s,\theta)/f_S(s) \\ &= \frac{\theta^{s+a-1}(1-\theta)^{(n-s)+b-1}}{\int_{\theta'}(\theta')^{s+a-1}(1-\theta')^{(n-s)+b-1}d\theta'} \\ &= \frac{\theta^{s+a-1}(1-\theta)^{(n-s)+b-1}}{Beta((s+a-1,(n-s)+b-1)}\end{aligned}$$

  This is a $Beta(a^*, b^*)$ distribution with

  $a^* = s + a$ and $b^* = (n-s) + b$.

Note:

- A prior distribution $Beta(a, b)$ corresponds to a prior belief consistent with hypothetical prior data consisting of
  - a successes and b failures, and
  - uniform "pre-hypothetical" prior.

Bayesian Inference: Normal Sample

Normal Sample

- $X_1, X_2, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$.
- Sample Space: $\mathcal{X} = (-\infty, +\infty)$ (for each X_i)
- Probability density function:
  $$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$
- Consider re-parametrization:
  $\xi = 1/\sigma^2$ (the **precision**) and $\theta = \mu$.
  $$f(x \mid \theta, \xi) = (\tfrac{\xi}{2\pi})^{\frac{1}{2}} e^{-\frac{1}{2}\xi(x-\theta)^2}$$

Three Cases:

- Unknown θ ($\xi = \xi_0$, known)
- Unknown ξ ($\theta = \theta_0$, known)
- Both θ and ξ unknown

Bayesian Inference: Normal Sample

Case 1: Unknown mean θ and known precision ξ_0

- Likelihood of sample $\mathbf{x} = (x_1, \ldots, x_n)$

$$\begin{aligned} lik(\theta) &= f(x_1, \ldots, x_n \mid \theta, \xi_0) \\ &= \prod_{i=1}^n f(x_i \mid \theta, \xi_0) \\ &= \prod_{i=1}^n \left(\frac{\xi_0}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}\xi_0 (x_i - \theta)^2} \\ &= \left(\frac{\xi_0}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\xi_0 \sum_{i=1}^n (x_i - \theta)^2} \end{aligned}$$

- Prior distribution: $\theta \sim N(\theta_0, \xi_{prior}^{-1})$

$$\pi(\theta) = \left(\frac{\xi_{prior}}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}\xi_{prior}(\theta - \theta_0)^2}$$

- Posterior distribution

$$\begin{aligned} \pi(\theta \mid \mathbf{x}) &\propto lik(\theta) \times \pi(\theta) \\ &= \left(\frac{\xi_0}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\xi_0 \sum_{i=1}^n (x_i - \theta)^2} \times \left(\frac{\xi_{prior}}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}\xi_{prior}(\theta - \theta_0)^2} \\ &\propto e^{-\frac{1}{2}[\xi_0 \sum_{i=1}^n (x_i - \theta)^2 + \xi_{prior}(\theta - \theta_0)^2]} \\ &\propto e^{-\frac{1}{2}[\xi_0 n(\theta - \overline{x})^2 + \xi_{prior}(\theta - \theta_0)^2]} \end{aligned}$$

(all constant factor terms dropped)

Bayesian Inference: Normal Sample

Case 1: Unknown mean θ and known precision ξ_0

- Claim: posterior distribution is Normal(!)

  Proof:

$$\begin{aligned} \pi(\theta \mid \mathbf{x}) &\propto lik(\theta) \times \pi(\theta) \\ &\propto e^{-\frac{1}{2}[\xi_0 n(\theta - \overline{x})^2 + \xi_{prior}(\theta-\theta_0)^2]} \\ &\propto e^{-\frac{1}{2}Q(\theta)} \end{aligned}$$

  where

$$Q(\theta) = \xi_{post}(\theta - \theta_{post})^2$$

  with

$$\begin{aligned} \xi_{post} &= \xi_{prior} + n\xi_0 \\ \theta_{post} &= \frac{(\xi_{prior})\theta_0 + (n\xi_0)\overline{x}}{(\xi_{prior}) + (n\xi_0)} \\ &= \alpha\theta_0 + (1-\alpha)\overline{x}, \quad \text{where } \alpha = \xi_{prior}/\xi_{post} \end{aligned}$$

  By examination: $\theta \mid \mathbf{x} \sim N(\theta_{post}, \xi_{post}^{-1})$

  Note: As $\xi_{prior} \longrightarrow 0$, $\quad \theta_{post} \longrightarrow \overline{x} = \hat{\theta}_{MLE}$

$$\xi_{post} \longrightarrow n\xi_0 \quad (\sigma^2_{post} \longrightarrow \sigma_0^2/n)$$

Bayesian Inference: Normal Sample

Case2: Unknown precision ξ and known mean θ_0.

- Likelihood of sample $\mathbf{x} = (x_1, \ldots, x_n)$

$$\begin{aligned} lik(\xi) &= f(x_1, \ldots, x_n \mid \theta_0, \xi) \\ &= \textstyle\prod_{i=1}^{n} f(x_i \mid \theta_0, \xi) \\ &= \textstyle\prod_{i=1}^{n} \left(\frac{\xi}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}\xi(x_i-\theta_0)^2} \\ &= \left(\frac{\xi}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\xi \quad {}_{i=1}^{n}(x_i-\theta_0)^2} \end{aligned}$$

- Prior distribution: $\xi \sim Gamma(\alpha, \lambda)$

$$\pi(\xi) = \frac{\lambda^{\alpha}\xi^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda\xi},\ \xi > 0 \text{ (}\textbf{“Conjugate” Prior}\text{)}$$

- Posterior distribution

$$\begin{aligned} \pi(\xi \mid \mathbf{x}) &\propto lik(\xi) \times \pi(\xi) \\ &= \left[\left(\frac{\xi}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\xi\sum_{i=1}^{n}(x_i-\theta_0)^2}\right] \times \left[\frac{\lambda^{\alpha}\xi^{-(\alpha-1}}{\Gamma(\alpha)} e^{-\lambda\xi}\right] \\ &\propto \xi^{\frac{n}{2}+\alpha-1} e^{-(\lambda+\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta_0)^2)\xi} = \xi^{\alpha^*-1} e^{-\lambda^*\xi} \end{aligned}$$

  $Gamma(\alpha^*, \lambda^*)$ distribution density with

$$\alpha^* = \alpha + \tfrac{n}{2} \text{ and } \lambda^* = \lambda + \tfrac{1}{2}\textstyle\sum_{i=1}^{n}(x_i - \theta_0)^2.$$

Bayesian Inference: Normal Sample

Case2: Unknown precision ξ and known mean θ_0 (continued)

- Posterior distribution

$$\begin{aligned}\pi(\xi \mid \mathbf{x}) &\propto lik(\xi) \times \pi(\xi) \\ &\propto \xi^{\frac{n}{2}+\alpha-1} e^{-(\lambda+\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta_0)^2)\xi} = \xi^{\alpha^*-1} e^{-\lambda^*\xi}\end{aligned}$$

  $Gamma(\alpha^*, \lambda^*)$ distribution density with

$$\alpha^* = \alpha + \frac{n}{2} \text{ and } \lambda^* = \lambda + \frac{1}{2}\sum_{i=1}^{n}(x_i - \theta_0)^2.$$

- Posterior mean: $E[\xi \mid \mathbf{x}] = \frac{\alpha^*}{\lambda^*}$
- Posterior mode: $mode(\pi(\xi \mid \mathbf{x})) = \frac{\alpha^*-1}{\lambda^*}$
- For small α and λ ,

$$\begin{aligned}E[\xi \mid \mathbf{x}] &\longrightarrow \frac{n}{\sum_{i=1}^{n}(x_i-\theta_0)^2} = 1/\hat{\sigma}^2_{MLE} \\ mode(\pi(\xi \mid \mathbf{x})) &\longrightarrow \frac{n-2}{\sum_{i=1}^{n}(x_i-\theta_0)^2} = (1-\frac{2}{n})/\hat{\sigma}^2_{MLE}\end{aligned}$$

Bayesian Inference: Normal Sample

Case3: Unknown mean θ and unknown precision ξ

- Likelihood of sample $\mathbf{x} = (x_1, \dots, x_n)$

$$\begin{aligned} lik(\theta, \xi) &= f(x_1, \dots, x_n \mid \theta, \xi) \\ &= \textstyle\prod_{i=1}^{n} f(x_i \mid \theta, \xi) \\ &= \textstyle\prod_{i=1}^{n} \left(\frac{\xi}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}\xi(x_i-\theta)^2} \\ &= \left(\frac{\xi}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\xi \quad {}_{i=1}^{n}(x_i-\theta)^2} \end{aligned}$$

- Prior distribution: θ and ξ independent, a priori with

$$\theta \sim N(\theta_0, {\xi_{prior}}^{-1})$$
$$\xi \sim Gamma(\alpha, \lambda)$$
$$\begin{aligned} \pi(\theta, \xi) &= \pi(\theta)\pi(\xi) \\ &= \left[\left(\frac{\xi_{prior}}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{1}{2}\xi_{prior}(\theta-\theta_0)^2}\right] \times \left[\frac{\lambda^{\alpha}\xi^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda\xi}\right] \end{aligned}$$

- Posterior distribution

$$\begin{aligned} \pi(\theta, \xi \mid \mathbf{x}) &\propto lik(\theta, \xi) \times \pi(\theta, \xi) \\ &\propto \left[(\xi)^{\frac{n}{2}} e^{-\frac{1}{2}\xi \sum_{i=1}^{n}(x_i-\theta)^2}\right] \times \left[e^{-\frac{1}{2}\xi_{prior}(\theta-\theta_0)^2}\right] \\ &\quad \times \left[\xi^{\alpha-1} e^{-\lambda\xi}\right] \end{aligned}$$

Bayesian Inference: Normal Sample Case 3

- Posterior distribution

$$\begin{aligned}\pi(\theta,\xi \mid \mathbf{x}) &\propto lik(\theta,\xi)\times\pi(\theta,\xi)\\ &\propto \left[(\xi)^{\frac{n}{2}}e^{-\frac{1}{2}\xi \quad {}_{i=1}^{n}(x_i-\theta)^2}\right]\times\left[e^{-\frac{1}{2}\xi_{prior}(\theta-\theta_0)^2}\right]\\ &\quad \times\ \xi^{\alpha-1}e^{-\lambda\xi}\end{aligned}$$

- Marginal Posterior distribution of θ :

$$\begin{aligned}\pi(\theta \mid \mathbf{x}) &= \int_{\{\xi\}}\pi(\theta,\xi \mid \mathbf{x})d\xi\\ &\propto \left[e^{-\frac{1}{2}\xi_{prior}(\theta-\theta_0)^2}\right]\times\int_{\{\xi\}}\left[(\xi)^{\alpha^*-1}e^{-\lambda^*\xi}\right]d\xi\\ &= \left[e^{-\frac{1}{2}\xi_{prior}(\theta-\theta_0)^2}\right]\times\frac{\Gamma(\alpha^*)}{(\lambda^*)^{\alpha^*}}\end{aligned}$$

  where $\alpha^* = \alpha+\frac{n}{2}$ and $\lambda^* = \lambda+\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2$.

- Limiting case as ξ_{prior}, α and $\lambda \longrightarrow 0$

$$\begin{aligned}\pi(\theta \mid \mathbf{x}) &\propto (\lambda^*)^{-\alpha^*} = \left[\sum_{i=1}^{n}(x_i-\theta)^2\right]^{-\frac{n}{2}}\\ &= \left[(n-1)s^2+n(\theta-\overline{x})^2\right]^{-\frac{n}{2}}\\ &\propto \left[1+\frac{1}{n-1}\frac{n(\theta-\overline{x})^2}{s^2}\right]^{-\frac{n}{2}}\end{aligned}$$

  Note: A posteriori $\sqrt{n}(\theta-\overline{x})/s \sim t_{n-1}$ (for small $\xi_{prior}, \alpha, \lambda$)

Bayesian Inference: Poisson Distribution

Poisson Sample

- $X_1, X_2, \ldots, X_n$ i.i.d. $Poisson(\lambda)$
- Sample Space: $\mathcal{X} = \{0, 1, 2, \ldots\}$ (for each X_i)
- Probability mass function:
$$f(x \mid \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$
- Likelihood of sample $\mathbf{x} = (x_1, \ldots, x_n)$
$$\begin{aligned} lik(\lambda) &= f(x_1, \ldots, x_n \mid \lambda) \\ &= \prod_{i=1}^n f(x_i \mid \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \\ &\propto \lambda^{\sum_1^n x_i} e^{-n\lambda} \end{aligned}$$
- Prior distribution: $\lambda \sim Gamma(\alpha, \nu)$
$$\pi(\lambda) = \frac{\nu^\alpha \lambda^{\alpha-1}}{\Gamma(\alpha)} e^{-\nu\lambda}, \quad \lambda > 0$$
- Posterior distribution
$$\begin{aligned} \pi(\lambda \mid \mathbf{x}) &\propto lik(\lambda) \times \pi(\lambda) = \left[\lambda^{\sum_1^n x_i} e^{-n\lambda}\right] \times \left[\frac{\nu^\alpha \lambda^{\alpha-1}}{\Gamma(\alpha)} e^{-\nu\lambda}\right] \\ &\propto \lambda^{\alpha^* - 1} e^{-\nu^* \lambda} \end{aligned}$$

$Gamma(\alpha^*, \nu^*)$ with $\alpha^* = \alpha + \sum_1^n x_i$ and $\nu^* = \nu + n$.

Bayesian Inference: Poisson Distribution

Specifying the prior distribution: $\lambda \sim Gamma(\alpha, \nu)$.

- Choose α and ν to match prior mean and prior variance

$$E[\lambda \mid \alpha, \nu] = \alpha/\nu \quad (= \mu_1)$$
$$Var[\lambda \mid \alpha, \nu] = \alpha/\nu^2 \quad (= \sigma^2 = \mu_2 - \mu_1^2)$$

  Set $\nu = \mu_1/\sigma^2$ and $\alpha = \mu_1 \times \nu$

- Consider uniform distribution on interval

$$[0, \lambda_{MAX}] = \{\lambda : 0 < \lambda < \lambda_{MAX}\}$$

  (Choose λ_{MAX} to be very large)

Example 8.4.A Counts of asbestos fibers on filters (Steel et al. 1980).

- 23 grid squares with mean count: $\overline{x} = \frac{1}{23}\sum_{i=1}^{2} 3x_i = 24.9$.

$$\hat{\lambda}_{MOM} = \hat{\lambda}_{MLE} = 24.9$$
$$StError(\hat{\lambda}) = \sqrt{\widehat{Var}(\hat{\lambda})} = \sqrt{\hat{\lambda}/n} = \sqrt{24.9/23} = 1.04$$

- Compare with Bayesian Inference ($\mu_1 = 15$ and $\sigma^2 = 5^2$)

Bayesian Inference: Hardy-Weinberg Model

Example 8.5.1 A / 8.6 C Multinomial sample

- Data: counts of multinomial cells, $(X_1, X_2, X_3) = (342, 500, 187)$, for $n = 1029$ outcomes corresponding to genotypes AA, Aa and aa which occur with probabilities: $(1-\theta)^2$, $2\theta(1-\theta)$ and θ^2.
- Prior for θ : Uniform distribution on $(0, 1) = \{\theta : 0 < \theta < 1\}$.
- Bayes predictive interval for θ agrees with approximate confidence interval based on $\hat{\theta} = 0.4247$.

See R Script implementing the Bayesian computations.

Bayesian Inference: Prior Distributions

Important Concepts

- **Conjugate Prior Distribution**: a prior distribution from a distriibution family for which the posterior distribution is from the same distribution family
    - Beta distributions for Bernoulli/Binomial Samples
    - Gamma distributions for Poisson Samples
    - Normal distributions for Normal Samples (unknown mean, known variance)
- **Non-informative Prior Distributions:** Prior distributions that let the data dominate the structure of the posterior distribution.
    - Uniform/Flat prior
    - Complicated by choice of scale/units for parameter
    - Non-informative prior density may not integrate to 1
      I.e., prior distribution is **improper**
    - Posterior distribution for improper priors corresponds to limiting case of sequence of proper priors.

Bayesian Inference: Normal Approximation to Posterior

Posterior Distribution With Large-Samples

- Conditional density/pmf of data: $X \sim f(x \mid \theta)$
- Prior density of parameter: $\theta \sim \pi(\theta)$
- Posterior density
$$\begin{aligned}\pi(\theta \mid x) &\propto \pi(\theta) f(x \mid \theta) \\ &= exp\left[\log \pi(\theta)\right] exp\left[\log f(x \mid \theta)\right] \\ &= exp\left[\log \pi(\theta)\right] exp\left[\ell(\theta)\right]\end{aligned}$$
- For a large sample, $\ell(\theta)$ can be expressed as a Taylor Series about the MLE $\hat{\theta}$
$$\begin{aligned}\ell(\theta) &= \ell(\hat{\theta}) + (\theta - \hat{\theta})\ell'(\hat{\theta}) + \tfrac{1}{2}(\theta - \hat{\theta})^2 \ell''(\hat{\theta}) \\ &\propto (\theta - \hat{\theta}) \cdot 0 + \tfrac{1}{2}(\theta - \hat{\theta})^2 \ell''(\hat{\theta}) \\ &= \tfrac{1}{2}(\theta - \hat{\theta})^2 \ell''(\hat{\theta})\end{aligned}$$
(i.e. Normal log-likelihood, mean $\hat{\theta}$ and variance $[\ell(\hat{\theta})]^{-1}$)
- For large sample, $\pi(\theta)$ is relatively flat in range near $\theta \approx \hat{\theta}$ and likelihood concentrates in same range.

18.443 Statistics for Applications
Spring 2015