Aims

- Normal model, mean unknown
- Predictive distribution
- Normal model, multiple obs., mean unknown
- Normal model, multiple obs., variance unknown
- Standard distributions

  Binomial

  Poisson (with application)

  Exponential
- Non informative Prior Distributions

  Jeffrey's invariance principle

Normal model, mean unknown, variance known

$$
\begin{aligned}
y &\sim N(\theta, \sigma^2), \\
p(y \mid \theta) &\propto \exp\{-\tfrac{1}{2\sigma^2}(y-\theta)^2\} \\
p(\theta) &\propto \exp\{-\tfrac{1}{2\tau_0^2}(\theta-\mu_0)^2\} \\
p(\theta \mid y) &\propto \exp\{-\tfrac{1}{2\tau_1^2}(\theta-\mu_1)^2\}
\end{aligned}
$$

1. $\mu_1 = \frac{\mu_0/\tau_0^2 + y/\sigma^2}{1/\tau_0^2 + 1/\sigma^2}, \quad \tau_1^2 = \frac{1}{1/\tau_0^2 + 1/\sigma^2}$
2. $\mu_1 = \mu_0 + (y - \mu_0)\frac{\tau_0^2}{\tau_0^2 + \sigma^2}$
3. $\mu_1 = y - (y - \mu_0)\frac{\sigma^2}{\tau_0^2 + \sigma^2}$

- $\mu_1 = \mu_0$ if $\tau_0^2 = 0$ or $y = \mu_0$
- $\mu_1 = y$ if $\sigma^2 = 0$ or $y = \mu_0$


```
normalnormal_function(w=0,y=2,sigmasqr=1,mu=seq(-2,6,.001)){
if(w == 1) postscript("/users/faculty/fdominic/teaching/BM/normalnormal.ps")
  par(mfrow=c(3,2))
  mu0_seq(0,5,1)
  tau0sqr_1.5
  tau1sqr_(1/sigmasqr+1/tau0sqr)^{-1}
  mu1_(y/sigmasqr+mu0/tau0sqr)*tau1sqr
  for(i in 1:length(mu0)){
    plot(mu,dnorm(mu,mu0[i],tau0sqr),xlab="mu",ylab="",
         xaxs="i",yaxs="i",yaxt="n",bty="n",xlim=c(-2,10),ylim=c(0,1))
    lines(mu,dnorm(mu,y,sigmasqr),lty=1)
    lines(mu,dnorm(mu,mu1[i],tau1sqr),lty=2)
  }
  par(oma=c(0,0,0,0))
  par(mfrow=c(1,1))
  if (w == 1) dev.off()}
```


Predictive Distribution

$\tilde{y}$ is a future observation

Prior Predictive Distribution

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

Posterior Predictive Distribution

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

$$
\begin{aligned}
E(\tilde{y} \mid y) &= E(E(\tilde{y} \mid \theta, y) \mid y) = E(\theta \mid y) = \mu_1 \\
Var(\tilde{y} \mid y) &= E(Var(\tilde{y} \mid \theta, y) \mid y) \\
&+ Var(E(\tilde{y} \mid \theta, y) \mid y) \\
&= E[\sigma^2 \mid y] + V[\theta \mid y)] \\
&= \sigma^2 + \tau_1^2
\end{aligned}
$$

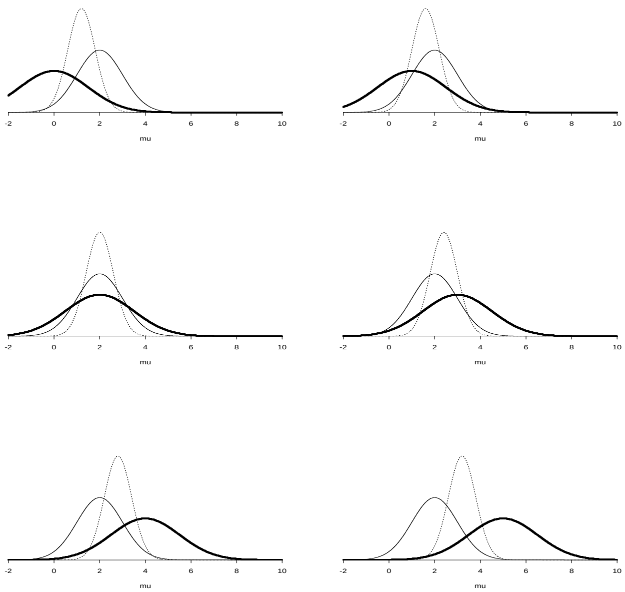


Figure 1: Likelihood, prior, posterior for μ.

Normal model with multiple observations, mean unknown, variance known

$$
\begin{aligned}
y_1, \ldots, y_n \mid \mu, \sigma^2 &\sim N(\theta, \sigma^2), \text{ iid} \\
p(y \mid \theta) &\propto \textstyle\prod_{i=1}^n \exp\{-\frac{1}{2\sigma^2}(y_i - \theta)^2\} \\
p(\theta \mid y_1, \ldots, y_n) &\propto N(\theta \mid \mu_n, \tau_n^2)
\end{aligned}
$$

- $\mu_n = \frac{\mu_0/\tau_0^2 + n\bar{y}/\sigma^2}{1/\tau_0^2 + n/\sigma^2}$
- $\tau_n^2 = \frac{1}{1/\tau_0^2 + n/\sigma^2}$

If $\tau_0 \to \infty$ and n fixed or $n \to \infty$ and τ_0 fixed then

$$\theta \mid y \sim N(\theta \mid \bar{y}, \sigma^2/n)$$

If the prior is diffuse then the posterior is a good approximation of the likelihood

Standard Distributions

- Binomial : counting exchangeable outcomes
- Normal: $X = \sum_{i=1}^N Z_i$, N large, Z_i independent
- Poisson: number of occurences in a given interval
- Exponential: waiting time

For all these cases, closed form of the normalizing constant $p(y)$ are available.

Normal model, variance unknown, mean known

$$
\begin{aligned}
p(y \mid \sigma^2) &\propto {\sigma^2}^{-n/2} \exp\{-\tfrac{1}{2\sigma^2}\textstyle\sum(y_i - \theta)^2\} \\
p(\sigma^2) &\propto {\sigma^2}^{-(\alpha+1)} \exp(-\tfrac{\beta}{\sigma^2}) \\
p(\sigma^2 \mid y) &\propto IG(\tfrac{n}{2} + \alpha, \beta + \tfrac{v}{2}),\ v = \tfrac{1}{2}\textstyle\sum_i^n (y_i - \theta)^2
\end{aligned}
$$

A convenient reparametrization is

- $\alpha = \frac{\nu_0}{2}$
- $\beta = \sigma_0^2 \frac{\nu_0}{2}$
- $\nu_0 = 0$ non informative prior

in that case...

$$
\begin{aligned}
p(\sigma^2) &\propto \chi^2_{inv}(\nu_0, \sigma_0^2) \\
p(\sigma^2 \mid y) &\propto \chi^2_{inv}(\nu_0 + n, \tfrac{\nu_0\sigma_0^2 + nv}{\nu_0 + n})
\end{aligned}
$$

NB: $\chi^2_{inv}(\nu) = IG(\frac{\nu}{2}, \frac{1}{2})$

Poisson Model

y is a count, for ex. an incidence of disease

$$y_1, \ldots, y_n \sim \text{Poisson}(y \mid \theta)$$

$$p(y \mid \theta) \propto \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!}$$

$$\propto \theta^{\sum y_i} e^{-n\theta}$$

$$p(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta} \propto \text{Gamma}(\theta \mid, \alpha, \beta)$$

$$p(\theta \mid y) \propto \theta^{\alpha-1+\sum y_i} e^{-\theta(\beta+n)}$$

$$\propto \text{Gamma}(\alpha + \textstyle\sum y_i, \beta + n)$$

Negative Binomial

The prior predictive of a Poisson model has a negative-binomial distribution

$$p(y) = \frac{\text{Poisson}(y|\theta)\text{Gamma}(\theta|\alpha,\beta)}{\text{Gamma}(\theta|\alpha+y,1+\beta)}$$

$$p(y) = \binom{\alpha+y-1}{y} \left(\frac{\beta}{\beta+1}\right)^{\alpha} \left(\frac{1}{\beta+1}\right)^{y}$$

$$= \text{Neg-Bin}(y \mid \alpha, \beta)$$

$$\text{Neg-Bin}(y \mid \alpha, \beta) = \int \text{Poi}(y \mid \theta)\text{Gam}(\theta \mid \alpha, \beta) d\theta$$

Poisson model with exposure

- y_i is an incidence of disease of i-th unit
- x_i is the exposure of i-th unit
- θ is the rate

$$y_i \sim \text{Poisson}(\theta x_i) \; i = 1, \ldots, n$$
$$\theta \sim \text{Gamma}(\alpha, \beta)$$

$$p(y \mid \theta) \propto \theta^{(\sum_i y_i)} e^{-\theta(\sum_i x_i)}$$

$$\theta \mid y \sim \text{Gamma}(\alpha + \textstyle\sum y_i, \beta + \sum x_i)$$

Estimating a rate from Poisson data

- In a single year, it is found that 3 persons, out of the population of 200,000 died of a asthma
- $y = 3$ is the number of deaths in a city of 200,000 persons in one year
- $x = 2$ exposure
- θ true underlying long-term asthma mortality rate (measured in cases per 100,000 persons per year)
- $y \sim \text{Poisson}(2\theta)$
- $\theta \sim \text{Gamma}(3, 5)$
- $\theta \mid y \sim \text{Gamma}(6, 7)$

```
#example pag 50 gelman book: estimating a rate from Poisson data:
#an idealized example
exposure_function(w=0){
  if (w == 1) postscript("/users/faculty/fdominic/teaching/BM/exposure.ps")
  par(mfrow=c(2,1))
  prior.theta_rgamma(1000,3)/5
  hist(prior.theta,nclass=50,xlab="theta",ylab="prior",yaxt="n",
       density=-1,prob=T,cex=2 ,xlim=c(0,3))
  abline(v=mean(prior.theta),lwd=2)
  mtext("PRIOR",side=3)
  posterior.theta_rgamma(1000,6)/7
  hist(posterior.theta,nclass=50,xlab="theta",ylab="posterior",yaxt="n",
       density=-1,prob=T,cex=2,xlim=c(0,3))
  abline(v=mean(posterior.theta),lwd=2)
  mtext("POSTERIOR",side=3)
  ppp_ sum(posterior.theta>1)/1000
  par(oma=c(0,0,0,0))
  par(mfrow=c(1,1))
  if (w == 1) dev.off()
}
```


Exponential model

- y is a waiting time
- Memoryless property:

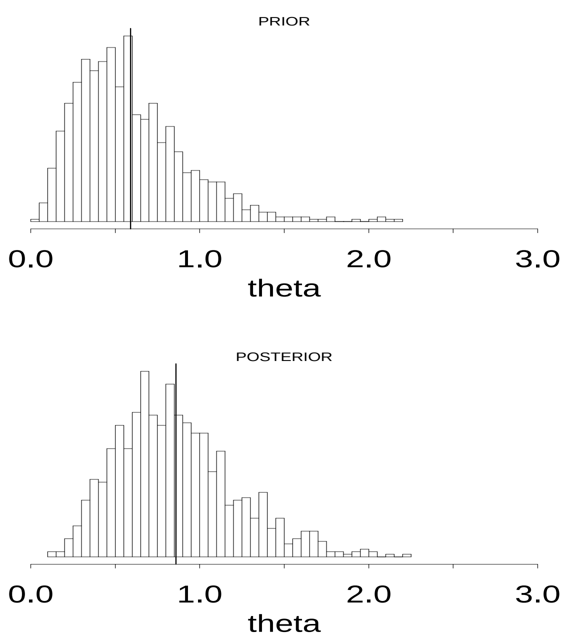
$$P(y > t + s \mid y > s, \theta) = P(y > t \mid \theta) \quad \forall t, s$$

$$p(y \mid \theta) \propto \theta \exp(-y\theta)$$

$$p(\theta) \propto \text{Gamma}(\theta \mid \alpha, \beta)$$

$$p(\theta \mid y) \propto \text{Gamma}(\theta \mid \alpha + 1, \beta + y)$$

- $\text{Exp}(\theta) = \text{Gamma}(\alpha = 1, \beta = \theta)$



Non informative prior distributions

- Reference prior: *let the data speak for themselves*
- Proper prior: it integrates to any positive number

Example: mean known, variance unknown

$$y \mid \theta \sim N(\theta, \sigma^2)$$
$$\theta \sim N(\mu_0, \tau_0^2)$$
$$\theta \mid y \sim N\left(\theta \mid \frac{\mu_0/\tau_0^2 + n\bar{y}/\sigma^2}{1/\tau_0^2 + n/\sigma^2}, \frac{1}{1/\tau_0^2 + n/\sigma^2}\right)$$

If $\tau_0 = \infty$ and n fixed or $n = \infty$ and τ_0 fixed then

$$\theta \sim \text{constant}$$
$$\theta \mid y \sim N(\theta \mid \bar{y}, \sigma^2/n)$$

NB: Improper priors can lead to proper posteriors

Example: mean known, variance unknown

$$y \mid \theta \sim N(\theta, \sigma^2)$$
$$\sigma^2 \sim \mathsf{IG}(\alpha, \beta)$$

If $\alpha = \beta = 0$ then

$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$
$$\sigma^2 \mid y \sim \mathsf{IG}(\tfrac{n}{2}, \tfrac{1}{2}\sum_{i=1}^{n}(y_i - \mu)^2)$$

NB: Improper priors can lead to proper posteriors

Jeffrey's Invariance Principle

Any rule for determining prior density for θ should yield an equivalent result if applied to a transformed parameter $\phi = h(\theta)$

$$p(\phi) = p(\theta)\left|\frac{d\theta}{d\phi}\right| = p(\theta)\left|h'(\theta)\right|^{-1}$$

Jeffrey's choice for non informative prior density is

$$p(\theta) \propto J(\theta)^{1/2}$$

where $J(\theta)$ is the Fisher information for θ

$$J(\theta) = -E\left[\frac{d^2 \log p(y \mid \theta)}{d\theta^2} \mid \theta\right]$$

In fact

$$J(\phi)^{1/2} = J(\theta)^{1/2}\left|\frac{d\theta}{d\phi}\right|$$

Various non-informative prior for the binomial parameter

- $y \mid \theta \sim$ Binomial$(y \mid \theta, n)$
- $J(\theta) = \frac{n}{\theta(1-\theta)}$
- Jeffrey's prior is
$$p(\theta) \propto \frac{1}{[\theta(1-\theta)]^{1/2}} \sim \text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$$
- Bayes-Laplace prior is uniform on $[0, 1]$
$$p(\theta) \sim \text{Beta}\ (\theta \mid 1, 1)$$
- Prior density uniform on the logit(θ)
$$p(\text{logit}\ \theta) \propto \text{cost then Beta}\ (\theta \mid 0, 0)$$

CAREFUL... If we use $\theta \sim$ Beta$(0, 0)$ and $y = 0$, then the posterior is improper!!

Difficulties with non informative

- if the likelihood is truly dominant , then the choice among a range of relatively flat prior densities cannot matter.
- A density that is flat in one parametrization will not be in another, ex. $p(\log \sigma^2) = 1$ but $p(\sigma^2) = \frac{1}{\sigma^2}$

Advantages with non informative

- it does not seem to be worth the effort to quantify one's real prior knowledge as a probability distribution, as long as one is willing to perform the mathematical work to check that the posterior density is proper and to determine sensitivity analysis.