Introduction to Bayesian Statistics
Lecture 3: Single Parameter (II)

Rung-Ching Tsai

Department of Mathematics
National Taiwan Normal University

March 11, 2015

Conjugate Prior Distributions

Definition of Conjugacy: If $\mathcal{F}$ is a class of sampling distributions $p(y|\theta)$, and $\mathcal{P}$ is a class of prior distributions for θ, then the class $\mathcal{P}$ is conjugate for $\mathcal{F}$ if

$$p(\theta|y) \in \mathcal{P} \text{ for all } p(\cdot|\theta) \in \mathcal{F} \text{ and } p(\theta) \in \mathcal{P}.$$

- Advantages of using conjugate priors:
  - computational convenience
  - being interpretable as additional data
- Example: Beta is conjugate for binomial with $\theta \sim \text{Beta}(\alpha, \beta)$ and $\theta|y \sim \text{Beta}(\alpha + y, \beta + n - y)$.
- Exercise: What is the conjugate prior for Poisson(λ)?

Conjugate Prior Distributions for exponential families

Definition: The class $\mathcal{F}$ is an exponential family if all its members have the form

$$p(y_i|\theta) = f(y_i)g(\theta)e^{\phi(\theta)^T u(y_i)},$$

where $\phi(\theta)$: the "natural parameter" of the family $\mathcal{F}$.

Exercise: Show that the binomial(n, θ) is an exponential family with natural parameter logit(θ), and the conjugate prior on θ are Beta distributions.

Conjugate Prior Distributions for exponential families

- Likelihood of θ:

$$\begin{aligned} p(\mathbf{y}|\theta) &= \left[\prod_{i=1}^{n} f(y_i)\right] g(\theta)^n \exp\left(\phi(\theta)^T \sum_{i=1}^{n} u(y_i))\right) \\ &\propto g(\theta)^n \exp\left(\phi(\theta)^T t(\mathbf{y})\right), \end{aligned}$$

  where $t(\mathbf{y}) = \sum_{i=1}^{n} u(y_i)$: sufficient statistic for θ
- (Conjugate) Prior:

$$p(\theta) \propto g(\theta)^\eta \exp\left(\phi(\theta)^T \nu\right)$$

- Posterior:

$$p(\theta|\mathbf{y}) \propto g(\theta)^{\eta+n} \exp\left(\phi(\theta)^T (\nu + t(\mathbf{y})\right).$$

- Known fact: Exponential families are, in general, the only classes of distributions that have natural conjugate priors.

Single Parameter θ: Continuous y

- $y \sim \text{normal}(\theta, \sigma^2)$, σ^2 known, use Bayesian approach to estimate θ.
  - choose a conjugate prior for θ, $p(\theta) = e^{A\theta^2 + B\theta + C}$, such that
$$p(\theta) \propto \exp\left(-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2\right)$$
  - likelihood of θ: $p(y|\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2\right)$
  - find the posterior distribution of θ:
$$\begin{aligned} p(\theta|y) \propto p(\theta)p(y|\theta) &\propto \exp\left(-\frac{1}{2}\left[\frac{(y-\theta)^2}{\sigma^2} + \frac{(\theta-\mu_0)^2}{\tau_0^2}\right]\right) \\ &\propto \exp\left(-\frac{1}{2\tau_1^2}(\theta - \mu_1)^2\right), \end{aligned}$$

  that is, $\theta|y \sim \text{normal}(\mu_1, \tau_1^2)$, where
  $\mu_1 = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{1}{\sigma^2}y}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}$ and $\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}$.

Single Parameter θ: Continuous y

- $\theta \sim \text{normal}(\mu_0, \tau_0^2), y \sim \text{normal}(\theta, \sigma^2) \Rightarrow \theta|y \sim \text{normal}(\mu_1, \tau_1^2)$
- posterior precision $\frac{1}{\tau_1^2}$
  - Definition of precision: the inverse of variance
  - $\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}$, i.e., the posterior precision equals the prior precision plus the data precision.
- posterior mean μ_1
  - $\mu_1 = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{1}{\sigma^2}y}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}$, i.e., the posterior mean is a weighted average of the prior mean and the observed value y, with weights proportional to the precision.
  - the prior mean adjusted toward the observed y: $\mu_1 = \mu_0 + (y - \mu_0)\frac{\tau_0^2}{\sigma^2 + \tau_0^2}$.
  - a compromise between the prior mean and the observed data y, with data shrunk toward the prior mean: $\mu_1 = y - (y - \mu_0)\frac{\sigma^2}{\sigma^2 + \tau_0^2}$

Single Parameter θ: Continuous y

Posterior predictive distribution $p(\tilde{y}|y)$

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta \\ &\propto \int \exp\left(-\frac{1}{2\sigma^2}(\tilde{y}-\theta)^2\right)\exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right)d\theta \end{aligned}$$

- $\tilde{y}|y \sim \text{normal}(?,?)$
- $\mathrm{E}(\tilde{y}|y) = \mathrm{E}(\mathrm{E}(\tilde{y}|\theta,y)|y) = \mathrm{E}(\theta|y) = \mu_1$
- $\mathrm{var}(\tilde{y}|y) = \mathrm{E}(\mathrm{var}(\tilde{y}|\theta,y)|y) + \mathrm{var}(\mathrm{E}(\tilde{y}|\theta,y)|y) = \mathrm{E}(\sigma^2|y) + \mathrm{var}(\theta|y) = \sigma^2 + \tau_1^2.$

Note. $\mathrm{E}(\tilde{y}|\theta) = \theta$, $\mathrm{var}(\tilde{y}|\theta) = \sigma^2$

Single Parameter θ: Continuous $\mathbf{y} = (y_1, \cdots, y_n)$

- $y_1, \cdots, y_n \overset{iid}{\sim}$ normal(θ, σ^2), σ^2 known, use Bayesian approach to estimate θ.
    - choose a conjugate prior for θ, $p(\theta) \propto \exp\left(-\frac{1}{2\tau_0^2}(\theta - \mu_0)^2\right)$
    - likelihood of θ: $p(\mathbf{y}|\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \theta)^2\right)$
    - find the posterior distribution of θ:

$$
\begin{aligned}
p(\theta|y) \propto p(\theta)p(\mathbf{y}|\theta) \quad & \propto \quad \exp\left(-\frac{1}{2}\left[\frac{\sum_{i=1}^n (y_i - \theta)^2}{\sigma^2} + \frac{(\theta - \mu_0)^2}{\tau_0^2}\right]\right) \\
& \propto \quad \exp\left(-\frac{1}{2\tau_n^2}(\theta - \mu_n)^2\right),
\end{aligned}
$$

that is, $\theta|\mathbf{y} \sim$ normal(μ_n, τ_n^2), where
$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$ and $\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$.

Single Parameter θ: Continuous $\mathbf{y} = (y_1, \cdots, y_n)$

- $y_1, \cdots, y_n \overset{iid}{\sim} \text{normal}(\theta, \sigma^2)$, σ^2 known, $\theta \sim \text{normal}(\mu_0, \tau_0^2)$
$\Rightarrow \theta|\mathbf{y} \sim \text{normal}(\mu_n, \tau_n^2)$
- posterior precision $\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$; posterior mean $\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$
    - If n is large, the posterior distribution is largely determined by σ^2 and the sample value $\bar{y}$.
    - As $\tau_0 \to \infty$ with n fixed, or as $n \to \infty$ with τ_0^2 fixed, we have

$$p(\theta|\mathbf{y}) \approx \text{normal}(\theta|\bar{y}, \frac{\sigma^2}{n}).$$

    - Compare the well-known result of classical statistics: $\bar{y}|\theta, \sigma^2 \sim \text{normal}(\theta, \frac{\sigma^2}{n})$ leads to the use of $\bar{y} \pm 1.96\frac{\sigma}{\sqrt{n}}$ as a 95% confidence interval for θ.
    - Bayesian approach gives the same result for noninformative prior.

Exercise

A random sample of n students is drawn from a large population, and their weights are measured. The average weight of the n sampled students is $\bar{y} = 150$ pounds. Assume the weights in the population are normally distributed with unknown mean θ and known standard deviation 20 pounds. Suppose your prior distribution for θ is normal with mean 180 and standard deviation 40.

(a) Give your posterior distribution for θ.

(b) A new student is sampled at random from the same population and has a weight of $\tilde{y}$ pounds. Give a posterior predictive distribution for $\tilde{y}$.

(c) For $n = 10$, give a 95% posterior interval for θ and a 95% posterior predictive interval for $\tilde{y}$.

(d) Do the same for $n = 100$.

Single Parameter σ^2: Continuous $\mathbf{y} = (y_1, \cdots, y_n)$

- $y_1, \cdots, y_n \overset{iid}{\sim}$ normal(θ, σ^2), θ known, use Bayesian approach to estimate σ^2.
    - likelihood of σ^2:

$$
\begin{aligned}
p(\mathbf{y}|\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \theta)^2\right) \\
&\propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}(y_i - \theta)^2\right) \\
&= (\sigma^2)^{-\frac{n}{2}} \exp(-\frac{n}{2\sigma^2} v)
\end{aligned}
$$

      where $v = \frac{1}{n}\sum_{i=1}^{n}(y_i - \theta)^2$
    - choose a conjugate prior for σ^2 (inverse-gamma):

$$p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} e^{-\frac{\beta}{\sigma^2}}$$

Single Parameter σ^2: Continuous $\mathbf{y} = (y_1, \cdots, y_n)$

- $y_1, \cdots, y_n \stackrel{iid}{\sim} \text{normal}(\theta, \sigma^2)$, θ known, estimate σ^2.
  - likelihood of σ^2: $p(\mathbf{y}|\sigma^2) = (\sigma^2)^{-\frac{n}{2}} \exp(-\frac{n}{2\sigma^2} v)^2$
  - choose a conjugate prior for σ^2 (inverse-gamma):
    $p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} e^{-\frac{\beta}{\sigma^2}}$, i.e., $\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$
    Note. A scaled inverse-χ^2 distribution with scale σ_0^2 and ν_0 degrees of freedom: $\frac{\sigma_0^2 \nu_0}{X} \sim \chi^2_{\nu_0}$, i.e., $X \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$
  - find the posterior distribution of σ^2:

$$
\begin{aligned}
p(\sigma^2) &\propto p(\sigma^2) p(\mathbf{y}|\sigma^2) \\
&\propto \left(\frac{\sigma_0^2}{\sigma^2}\right)^{\nu_0/2+1} \exp\left(-\frac{\sigma_0^2 \nu_0}{2\sigma^2}\right) \cdot (\sigma^2)^{-\frac{n}{2}} \exp(-\frac{n}{2}\frac{v}{\sigma^2}) \\
&\propto (\sigma^2)^{-((n+\nu_0)/2+1)} \exp\left(-\frac{1}{2\sigma^2}(\nu_0 \sigma_0^2 + nv)\right).
\end{aligned}
$$

that is, $\sigma^2|\mathbf{y} \sim \text{Inv-}\chi^2\left(\nu_0 + n, \frac{\nu_0\sigma_0^2 + nv}{\nu_0 + n}\right)$.

Homework II

1. The following Table gives the number of fatal accidents and deaths on scheduled airline flights per year over a ten-year period.

Year	Fatal accidents	Passenger death	Death rate	Year	Fatal accidents	Passenger death	Death rate
1976	24	734	0.19	1981	21	362	0.06
1977	25	516	0.12	1982	26	764	0.13
1978	31	754	0.15	1983	20	809	0.13
1979	31	877	0.16	1984	16	223	0.03
1980	22	814	0.14	1985	22	1066	0.15

   (a) Assume that the number of fatal accidents in each year are independent with a Poisson(θ) distribution. Set a prior distribution for θ and determine the posterior distribution based on the data from 1976 through 1985. Under this model, give a 95% predictive interval for the number of fatal accident in 1986. You can use normal approximation to the gamma and Poisson or compute using simulation.

   (b) Repeat (a) above, replacing 'fatal accidents' with 'passenger deaths'.

Homework II

2. Censored and uncensored data in the exponential model:
   (a) Suppose $y|\theta$ is exponentially distributed with rate θ, and the marginal (prior) distribution of θ is Gamma(α, β). Suppose we observe that $y \geq 100$, but do not observe the exact value of y. What is the posterior distribution, $p(\theta|y \geq 100)$, as a function of α and β? Write down the posterior mean and variance of θ.
   (b) In the above problem, suppose that we are now told that y is exactly 100. Now what are the posterior mean and variance of θ?