STA 630: Bayesian Inference - Chapter 3

Zeya Wang

University of Kentucky

Spring 2025

Lectures 5: Normal Model - Outline

- A multiparameter model: The normal model
- Non-informative, conjugate and semi-conjugate priors
- Example (Practice 3)
- Predictive distribution

The Normal Model

$x = (x_1, \ldots, x_n) \sim N\left(\mu, \sigma^2\right)$ i.i.d., with both μ and σ unknown. The likelihood is:

$$
\begin{aligned}
L\left(\mu, \sigma^2\right) &\propto \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left(x_i - \mu\right)^2\right) \\
&\propto \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_i \left(x_i - \mu\right)^2\right)
\end{aligned}
$$

For inference, focus is on $p\left(\mu, \sigma^2 \mid x\right) = p\left(\mu \mid \sigma^2, x\right) p\left(\sigma^2 \mid x\right)$. From a Bayesian perspective, it is easier to work with the precision, $\tau = \frac{1}{\sigma^2}$. The likelihood becomes:

$$
\begin{aligned}
L(\mu, \tau) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \tau^{1/2} \exp\left(-\frac{1}{2}\tau\left(x_i - \mu\right)^2\right) \\
&\propto \tau^{n/2} \exp\left(-\frac{1}{2}\tau\sum_i \left(x_i - \mu\right)^2\right)
\end{aligned}
$$

The Normal Model

Likelihood factorization:

$$
\begin{aligned}
L(\mu,\tau) &\propto \tau^{n/2} \exp\left(-\frac{1}{2}\tau \sum_i (x_i-\mu)^2\right) \\
&\propto \tau^{n/2} \exp\left(-\frac{1}{2}\tau \sum_i [(x_i-\bar{x})-(\mu-\bar{x})]^2\right) \\
&\propto \tau^{n/2} \exp\left(-\frac{1}{2}\tau \left[\sum_i (x_i-\bar{x})^2 + n(\mu-\bar{x})^2\right]\right) \\
&\propto \tau^{n/2} \exp\left(-\frac{1}{2}\tau s^2(n-1)\right) \exp\left(-\frac{1}{2}\tau n(\mu-\bar{x})^2\right) \\
&\propto \tau^{n/2} \exp\left(-\frac{1}{2}\tau SS\right) \exp\left(-\frac{1}{2}\tau n(\mu-\bar{x})^2\right)
\end{aligned}
$$

with $s^2 = \sum_i (x_i-\bar{x})^2/(n-1)$ and $SS = \sum_i (x_i-\bar{x})^2$ sample variance and sum of squares [SS and $\bar{x}$ sufficient statistics]

Non-informative Prior

Non-informative prior: $\pi\left(\mu,\sigma^2\right) \propto \frac{1}{\sigma^2}$. This arises by considering μ and σ^2 a priori independent and taking the product of the standard non-inf priors. This is not a conjugate setting (the posterior does not factor into a product of two independent distributions). Prior is improper but posterior is proper. This is also the Jeffreys' prior. Joint posterior distribution of μ and σ^2 is

$$p\left(\mu,\sigma^2 \mid x\right) \propto \left(\sigma^2\right)^{-(n/2+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]\right\}$$

where

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

Non-informative Prior

- The conditional posterior distribution, $p\left(\mu \mid \sigma^2, x\right)$, is equivalent to deriving the posterior for μ when σ^2 is known

$$\mu \mid \sigma^2, x \sim \mathcal{N}\left(\bar{x}, \frac{\sigma^2}{n}\right)$$

- The marginal posterior $p\left(\sigma^2 \mid x\right)$, is obtained integrating $p\left(\mu, \sigma^2 \mid x\right)$ over μ [Hint: integral of a Gaussian function $2c\sqrt{\pi} = \int 2\exp\left(-\frac{1}{c^2}(\mu+b)^2\right) d\mu$]

$$\begin{aligned} p\left(\sigma^2 \mid x\right) &\propto \int_\mu \left(\sigma^2\right)^{-(n/2+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]\right\} d\mu \\ &\propto \left(\sigma^2\right)^{-[(n-1)/2+1]} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\} \end{aligned}$$

which is an inverse-gamma density, i.e.

$$\sigma^2 \mid x \sim \text{ Inv-Gamma }\left(\frac{n-1}{2}, \frac{n-1}{2}s^2\right) \equiv \text{Inv-}\chi^2\left(n-1, s^2\right)$$

or, equivalently, $\tau \mid x \sim Ga$.

Sampling from the joint posterior distribution

One can simulate a value of (μ, σ^2) from the joint posterior density by (1) simulating σ^2 from an inverse-Gamma $\left(\frac{n-1}{2}, s^2\frac{n-1}{2}\right)$ distribution [take the inverse of random samples from a Gamma $\left(\frac{n-1}{2}, s^2\frac{n-1}{2}\right)$] (2) then simulating μ from $\mathcal{N}\left(\bar{x}, \frac{\sigma}{\sqrt{n}}\right)$ distribution.

Marginal posterior distribution $p(\mu \mid x)$ of μ

As μ is typically the parameter of interest (σ^2 nuisance parameter) it is useful to calculate its marginal posterior distribution [Hint: integral of a Gamma function $\frac{\Gamma(a)2^a}{b^a} = \int_0^\infty z^{a-1} \exp\left(-\frac{zb}{2}\right) dz$]

$$
\begin{aligned}
p(\mu \mid x) &= \int_0^\infty p\left(\mu, \sigma^2 \mid x\right) d\sigma^2 \\
&\propto \int_0^\infty \left(\sigma^2\right)^{-(n/2+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]\right\} d\sigma^2 \\
&= A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz, A = (n-1)s^2 + n(\bar{x}-\mu)^2, z = \frac{A}{2\sigma^2} \\
&\propto \quad A^{-n/2} = \left[1 + \frac{1}{n-1}\left(\frac{\mu-\bar{x}}{s/\sqrt{n}}\right)^2\right]^{-[(n-1)+1]/2}
\end{aligned}
$$

We recognize the kernel of the t-distribution, i.e., $\mu \mid x \sim t\left(n-1, \bar{x}, s^2/n\right)$, that is $\left.\frac{\mu-\bar{x}}{s/\sqrt{n}}\right| x \sim t_{n-1}$ with t_{n-1} the standard t-distribution with $n-1$ degrees of freedom (Note: t-distribution as a scale mixture of a Normal).

Conjugate Prior Model

A conjugate prior must be of the form $\pi\left(\mu, \sigma^2\right) = \pi\left(\mu \mid \sigma^2\right) \pi\left(\sigma^2\right)$, e.g.,

$$\mu \mid \sigma^2 \sim N\left(\mu_0, \sigma^2/\tau_0\right), \quad \sigma^2 \sim IG\left(\frac{\nu_0}{2}, \frac{SS_0^2}{2}\right) \quad [\text{ or } \tau \sim Ga]$$

which corresponds to the joint prior density

$$p\left(\mu, \sigma^2\right) \propto \left(\frac{\sigma^2}{\tau_0}\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2/\tau_0}\left(\mu - \mu_0\right)^2\right\} \left(\sigma^2\right)^{-(\nu_0/2+1)} \exp\left\{-\frac{SS_0^2}{2\sigma^2}\right\}$$

$$= \left(\sigma^2\right)^{-\left(\frac{\nu_0+1}{2}+1\right)} \exp\left\{-\frac{\tau_0}{2\sigma^2}\left(\frac{SS_0^2}{\tau_0} + \left(\mu - \mu_0\right)^2\right)\right\}$$

we call this a Normal-Inverse-Gamma prior,

$$\left(\mu, \sigma^2\right) \sim NIG\left(\mu_0, \tau_0, \nu_0/2, SS_0/2\right)$$

Joint Posterior $p(\mu, \sigma^2)$

$$
\begin{aligned}
p\left(\mu, \sigma^2 \mid x\right) \propto \left(\sigma^2\right)^{-\left(\frac{\nu_0+1}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}\left(SS_0^2 + \tau_0\left(\mu-\mu_0\right)^2\right)\right\} \\
\times \left(\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-\mu\right)^2\right\} \\
\propto \left(\sigma^2\right)^{-\left(\frac{\nu_n+1}{2}+1\right)} \exp\left\{-\frac{\tau_n}{2\sigma^2}\left(\frac{SS_n^2}{\tau_n} + \left(\mu-\mu_n\right)^2\right)\right\}
\end{aligned}
$$

with

$$
\mu \mid \sigma^2, x \sim N\left(\mu_n, \sigma^2/\tau_n\right), \quad \mu_n = \frac{\mu_0\frac{\tau_0}{\sigma^2} + \bar{x}\frac{n}{\sigma^2}}{\frac{\sigma_0}{\sigma^2} + \frac{n}{\sigma^2}} = \frac{\tau_0\mu_0 + n\bar{x}}{\tau_n}, \tau_n = \tau_0 + n
$$

$$
\sigma^2 \Big| x \sim IG\left(\frac{\nu_n}{2}, \frac{SS_n^2}{2}\right), \quad \nu_n = \nu_0 + n, SS_n = SS_0 + SS + \frac{\tau_0 n}{\tau_n}\left(\bar{x}-\mu_0\right)^2
$$

Thus, $\mu, \sigma^2 \mid x \sim$ Normal-Inverse Gamma $\left(\mu_n, \tau_n; \nu_n/2, SS_n^2/2\right)$.

Comments:

- μ_n expected value for μ after seeing the data

$$\mu_n = \frac{n}{\tau_n}\bar{x} + \frac{\tau_0}{\tau_n}\mu_0, \text{ weighted average}$$

- τ_n precision for estimating μ after n observations.
- ν_n degrees of freedom $\left[\tau \sim Ga(\alpha/2, \beta/2) \to \beta\tau \sim \chi^2_\alpha\right.$, with α degrees of freedom]
- SS_n posterior variation as prior variation+observed variation+variation between prior mean and sample mean.
- Limiting case $\tau_0 \to 0, \nu_0 \to -1($ and $SS_0 \to 0)$ then $\mu \mid x \sim t_{n-1}\left(\bar{x}, s^2/n\right)$ (same as improper prior!)

$$\text{Also } \mu \mid x \sim t_{\nu_n}\left(\mu_n, \sigma^2_n/\tau_n\right), \sigma^2_n = SS^2_n/\nu_n$$

[Note: Again $\int N\left(m, \sigma^2/\tau\right) Ga(\nu/2, SS/2) d\sigma^2 = t_\nu(m, SS/(\nu\tau))$]

Practice 3: Example on SPF (from Merlise Clyde)

A Sunlight Protection Factor (SPF) of 5 means an individual that can tolerate X minutes of sunlight without any sunscreen can tolerate $5X$ minutes with sunscreen. Data on 13 individual (tolerance, in min, with and without sunscreen). Analysis should take into account pairing which induces dependence between observations (take differences and use ratios or log (ratios) = difference in logs). Ratios make more sense given the goals: how much longer can a person be exposed to the sun relative to their baseline. Model: $Y = \log(TRT) - \log(\text{ CONTROL }) \sim N(\mu, \tau)$. Then $E(\log(\text{ TRT } / \text{ CONTROL })) = \mu = \log(SPF)$. Interested in $\exp(\mu) = SPF$. Summary statistics: $\bar{y} = 1.998, s^2 = 0.525, n = 13$ [make boxplots and Q-Q normal plots to check on normality]

Practice 3: Example on SPF (from Merlise Clyde)

Subjective prior for μ :

- Take prior median SPF to be 16
- $P(\mu > 64) = 0.01$
- information in prior is worth 25 observations

Solve for consistent hyperparameters: $\mu_0 = \log(16), \tau_0 = 25, \nu_0 = \tau_0 - 1$ and $P(\mu < \log(64)) = .99$ where $\frac{\mu - \mu_0}{\sqrt{SS_0/(\nu_0 \tau_0)}} \sim t_{\nu_0}$ (what is this?) implying $SS_0 = 185.7$.

Practice 3: Example on SPF (from Merlise Clyde)

Posterior hyperpar: $\tau_n = 38, \mu_n = 2.508, \nu_u = 37, SS_n = 197.134$
Sampling from posterior:
Draw $\tau \mid Y$

```
tau = rgamma(10000, vn/2, rate=SSn/2)
```

Draw $\mu \mid \tau, Y$

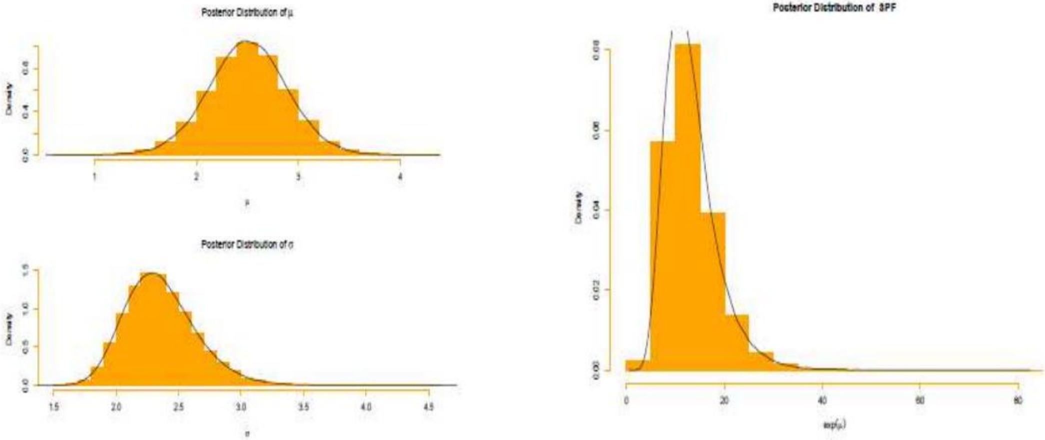
```
mu = rnorm(10000, mn, 1/sqrt(phi*pn))
```

or draw $\mu \mid Y$ directly

```
mu=rt(10000, v n) * sqrt(SSn /(pn * vn))+mn
```

Practice 3: Example on SPF (from Merlise Clyde)

Transform to $\exp(\mu)$. Find 95% C.I. of 4.54 to 23.758



Predictive Distribution of future z

- Posterior predictive distribution (given $x = (x_1, \ldots, x_n)$):

$$p(z \mid x) = \int p\left(z \mid \mu, \sigma^2, x\right) \pi\left(\mu, \sigma^2 \mid x\right) d\mu d\sigma^2$$

[Use assumption that z is independent of x given μ and σ^2, then integrate μ using the normal integral, then integrate σ^2 using the Gamma integral]

  - Reference prior: $z \mid x \sim t_{n-1}\left(\bar{x}, s^2(n+1)/n\right)$
  - Conjugate prior: $z \mid x \sim t_{\nu_n}\left(\mu_n, \sigma_n^2\left(\tau_n+1\right)/\tau_n\right)$

[Can use the normal "trick" to integrate μ : If $z \sim N\left(\mu, \sigma^2\right)$ and $\mu \sim N\left(\mu_0, \sigma^2/\tau_0\right)$ then $y = \frac{z-\mu}{\sigma} \sim N(0,1)$, that is $z = \sigma y + \mu$ and therefore $z \Big| \sigma^2 \sim N\left(\mu_0, \sigma^2\left(1+\frac{1}{\tau_0}\right)\right)$ since a linear comb of (independent) normals is normal with added mean and variance.]

- Prior predictive distribution: What we expect the distribution to be before we observe the data,

$$p(z) = \int p\left(z \mid \mu, \sigma^2\right) \pi\left(\mu, \sigma^2\right) d\mu d\sigma^2 \rightarrow z \sim t_{\nu_0}\left(\mu_0, \frac{SS_0}{\nu_0}\left(1 + \frac{1}{\tau_0}\right)\right)$$

[as above] $\left[\int N\left(\mu, \sigma^2\right) N\left(\mu_0, \sigma^2/\tau_0\right) Ga(\nu/2, SS/2) d\mu d\sigma^2 = t_\nu\left(\mu_0, \frac{SS}{\nu}\left(1 + \frac{1}{\tau_0}\right)\right)\right]$

Note: This is what we used in the example to specify our subjective prior.

Back to example

Prior predictive distribution: $z \sim t_{24}\left(\log(16), \frac{185.7}{24}\left(1+\frac{1}{25}\right)\right)$

Posterior predictive distribution: $z \sim t_{37}\left(2.5, 5.32\left(1+\frac{1}{38}\right)\right)$

```
Y=rt(10000,24) * sqrt((1+1 / 25) * 187.5 / 24)+log (16)
quantile(exp(Y))
0% 25% 50% 75% 100%
4.57 e-06    2.32.    16 .78     114 .98      370966 .2
```

Sampling from posterior predictive leads to 50% C.I. $(0.0003, 12.4)$ - with sunscreen, 50% chance that next individual can be exposed from 0 to 12 times longer than without sunscreen.

Semi-conjugate prior

A semi-conjugate setting is obtained with independent priors $\pi\left(\mu, \sigma^2\right) = \pi(\mu)\pi\left(\sigma^2\right)$

$$\mu \sim N\left(\mu_0, \sigma_0^2\right), \quad \sigma^2 \sim IG\left(\frac{\nu_0}{2}, \frac{\delta_0^2}{2}\right)$$

then $\mu \mid \sigma^2, x \sim N\left(\mu_n, \tau_n^2\right), \quad \mu_n = \frac{\frac{\mu_0}{\sigma_0^2} + \bar{x}\frac{n}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \quad \tau_n^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$

$\sigma^2 \mid x \sim$ not in closed form

We will solve this with MCMC methods!

Lecture 6: Outline

- Bayes factors
- The multinomial model
- Prelude to MCMC methods

Bayes Factors for Hypothesis Testing

- Bayes factors are used to test hypotheses and compare models in the Bayesian framework.
- Kass & Raftery (1995, JASA) is an excellent review
- Suppose we have two candidate models, M_1 and M_2, with respective parameter vectors θ_1 and θ_2.
- The Bayes factor in favor of M_1 is the ratio of the posterior odds of M_1 to the prior odds of M_1

$$BF = \frac{p\left(M_1 \mid x\right) / p\left(M_2 \mid x\right)}{p\left(M_1\right) / p\left(M_2\right)}$$

Bayes Factors for Hypothesis Testing

- The Bayes factor can also be written as the ratio of the observed marginal densities for the two models (via Bayes theorem)

$$BF = \frac{p\left(M_1 \mid x\right)/p\left(M_2 \mid x\right)}{p\left(M_1\right)/p\left(M_2\right)} = \frac{\left[\frac{p(x|M_1)p(M_1)}{p(x)}\right] \Big/ \left[\frac{p(x|M_2)p(M_2)}{p(x)}\right]}{p\left(M_1\right)/p\left(M_2\right)}$$
$$= \frac{p\left(x \mid M_1\right)}{p\left(x \mid M_2\right)}$$

- The marginal distribution of x under each model M_i is

$$p\left(x \mid M_i\right) = \int f\left(x \mid \theta_i, M_i\right) \pi_i\left(\theta_i \mid M_i\right) d\theta_i, \quad i = 1, 2$$

- In essence, how likely the data are, based on each model and integrating over the uncertainty in the parameters as represented by the prior.
- The Bayes factor is only defined when the marginal density of x under each model is proper. If $\pi_i\left(\theta_i\right)$ is improper, then $p\left(x \mid M_i\right)$ will necessarily be improper, and the Bayes factor is not defined

Interpretation of Bayes factor

Jeffreys' - scale of evidence in favor of M_1		
$\log_{10} BF$	Bayes factor	Interpretation
$0 - 0.5$	$1 \leq BF \leq 3.2$	weak
$0.5 - 1.0$	$3.2 < BF \leq 10$	substantial
$1.0 - 2.0$	$10 < BF \leq 100$	strong
> 2	$BF > 100$	decisive

Kass & Raftery — scale of evidence in favor of M_1		
$2 \ln BF$	Bayes factor	Interpretation
$0 - 2$	$1 \leq BF \leq 3$	weak
$2 - 6$	$3 < BF \leq 20$	positive
$6 - 10$	$20 < BF \leq 150$	strong
> 10	$BF > 150$	very strong

Comparison to frequentist hypothesis testing

In classical hypothesis testing, we proceed as follows:

1. state a null hypothesis, H_0, and an alternative hypothesis, H_1
2. determine an appropriate test statistic, $T(X)$
3. compute the p-value of the test as

$$p\text{-value} \;= P\left(T(X) \text{ more "extreme" than } T\left(x_{obs}\right) \mid \theta, H_0\right)$$

   where "extremeness" is in the direction of H_1

4. if the p-value is less than the prespecified Type I error rate, α, H_0 is rejected

Straightforward only when the two hypotheses are nested.

Comparison to frequentist hypothesis testing

In Bayesian hypothesis testing, we proceed as follows:

1. state the two hypotheses, M_1 and M_2
2. assign priors to M_1 and M_2, and specify $p\,(\theta \mid M_1)$ and $p\,(\theta \mid M_2)$
3. evaluate $P\,(M_1 \mid x)$ and $p\,(M_2 \mid x)$ via Bayes' theorem
4. compute the Bayes factor to assess the evidence in favor of M_1 :

$$BF = \frac{P\,(M_1 \mid x)\,/P\,(M_2 \mid x)}{P\,(M_1)\,/P\,(M_2)} = \frac{p\,(x \mid M_1)}{p\,(x \mid M_2)}.$$

Does not require the two models to be nested.

Example 1: Test of proportion

Suppose 16 customers have been recruited by a fast-food chain to compare two types of ground beef patty on the basis of flavor. All of the patties to be evaluated have been kept frozen for eight months.

- One set of 16 has been stored in a high-quality freezer that maintains a temperature that is consistently within $\pm 1^\circ \mathrm{F}$.
- The other set of 16 has been stored in a freezer with temperature that varies anywhere between 0 and $15^\circ F$.

The food chain executives are interested in whether storage in the higher-quality freezer translates into a substantial improvement in taste, thus justifying the extra effort and cost associated with equipping all of their stores with these freezers. Suppose that to be regarded as "substantial" improvement more than 60% of consumers must prefer the more expensive option. 13 of the 16 consumers state a preference for the more expensive patty.

Example 1: Test of proportion

- Let $Y_i = 1$ if consumer i states a preference for the more expensive patty and $Y_i = 0$ otherwise.

$$X = \sum_{i=1}^{16} Y_i \sim \text{Binomial}(16, \theta)$$

- We want to test:

$$M_1 : \theta > 0.6 \text{ vs } M_2 \leq 0.6$$

- Suppose we consider "minimally informative" priors, $\pi(\theta)$:
  - Jeffreys' prior, Beta$(.5, .5)$
  - a prior that we think of as "noninformative", Beta $(1, 1)$
  - Beta$(2, 2)$ prior

Example 1: Test of proportion

The posterior distribution for θ is given by

$$\theta \mid x \sim \text{Beta}(\alpha + x, \beta + n - x)$$

	Posterior quantile				
Prior	0.025	0.5	0.975	$p(\theta > 0.6 \mid x)$	BF
Beta $(.5, .5)$	0.579	0.806	0.944	0.964	34.432
Beta$(1, 1)$	0.566	0.788	0.932	0.954	30.812
Beta$(2, 2)$	0.544	0.758	0.909	0.930	24.604

Strong evidence in favor of $M_1 : \theta > 0.6$.

Example 2: Two-sided test of normal mean

John weighed 170 pounds last year and he is wondering if he still weighs the same. For simplicity, assume he knows the accuracy of the scale and $\sigma = 3$ pounds.

$$M_1 : \mu \neq 170 \quad \text{vs.} \quad M_2 : \mu = 170$$

He weighs himself 10 times and obtains the following measurements:

182 172 173 176 176 180 173 174 179 175

Example 2: Two-sided test of normal mean

- Under M_1, he may think that it's more likely that μ is close to 170 than far from it and can take normal distribution with mean 170 and standard deviation τ.
- Under $M_2, \pi(\mu) = I\{\mu = \mu_0\}$, a point mass at μ_0.
- The Bayes factor in support of M_1 is

$$BF = \frac{\left(\sigma^2/n + \tau^2\right)^{-1/2} \exp\left\{-\frac{1}{2(\sigma^2/n+\tau^2)} \left(\bar{x} - \mu_0\right)^2\right\}}{\sqrt{n}/\sigma \exp\left\{-\frac{n}{2\sigma^2} \left(\bar{x} - \mu_0\right)^2\right\}}$$

Example 2: Two-sided test of normal mean

Let us consider different values of τ.

```
weights = c(182,172,173,176,176,180,173,174,179,175)
xm = mean(weights); n = length(weights)
s=3; mu0 = 170; tau=c(0.5, 1, 2, 5, 10)
dnorm(xm,mu0,sqrt(s^2/n+tau^2))/dnorm(xm,mu0,s/sqrt(n))
[1]
6.8392e+01 2.5660e+04 5.2789e+06 4.5137e+07 3.8334e+07
```

Very strong evidence in favor of M_1; his current weight is substantially different from 170 lbs.

More comments

- BF reduces to the likelihood test for simple vs. simple hypothesis testing

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta = \theta_1$$

$$BF = \frac{p\left(x \mid \theta = \theta_0\right)}{p\left(x \mid \theta = \theta_1\right)}$$

- For general hypotheses the BF expressed in terms of the probability densities is

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_1, \quad \Theta = \Theta_0 \cup \Theta_1 ; \Theta_0 \cap \Theta_1 = \varnothing$$

$$BF = \frac{\int_{\theta \in \Theta_0} p\left(x \mid \theta, H_0\right) \pi\left(\theta \mid H_0\right) d\theta}{\int_{\theta \in \Theta_1} p\left(x \mid \theta, H_1\right) \pi\left(\theta \mid H_1\right) d\theta} = m_0(x)/m_1(x)$$

can be calculated directly or as ratio of posterior vs prior odds

- BF is defined ONLY for proper prior distributions and may be sensitive to prior choices, especially to weakly specifications, for example $\theta \sim N\left(\mu_0, \sigma_0^2\right)$ as $\sigma_0^2 \to \infty$.

More examples

- Simple vs. simple: $x = (x_1, \dots, x_n) \sim N(\mu, 1), H_0 : \mu = 0, \quad H_1 : \mu = 1$

$$BF = \frac{(2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_i x_i^2\right)}{(2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_i (x_i - 1)^2\right)} = \exp\left(\frac{n}{2} - \sum_i x_i\right)$$

If $n = 10, \sum_i x_i = 4.5$ then $BF = 1.65$ weak evidence in favor of H_0
If $n = 10, \sum_i x_i = 1$ then $BF = 55$ strong evidence in favor of H_0

- Simple vs. composite: $x = (x_1, \dots, x_n) \sim N(\mu, 1)$,

$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$

$$BF = \frac{p\left(x \mid \mu = \mu_0\right)}{\int_{\mu \in H_1} p(x \mid \mu)\pi(\mu) d\mu}$$

More examples

Assume $\mu_0 = 0$ and $\mu \sim N(1,1)$ then

$$BF = \frac{(2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_i x_i^2\right)}{(2\pi)^{-\frac{n}{2}} \int \exp(-\frac{1}{2}\sum_i (x_i - \mu)^2\Big) (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mu - 1)^2\right) d\mu} =$$

$$= \frac{(2\pi)^{1/2} \exp\left(-\frac{1}{2}\sum_i x_i^2\right)}{\int \exp\left(-\frac{1}{2}\left(\sum_i x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right) \exp\left(-\frac{1}{2}\left(\mu^2 - 2\mu + 1\right)\right) d\mu}$$

More examples

$$= \frac{(2\pi)^{1/2} \exp\left(-\frac{1}{2}\sum_i x_i^2\right)}{\int \exp\left(-\frac{1}{2}\left(1+\sum_i x_i^2\right)\right) \exp\left(\frac{\left(1+\sum x_i\right)^2}{2(n+1)}\right) \exp\left(-\frac{n+1}{2}\left(\mu - \frac{1+\sum x_i}{n+1}\right)^2\right) d\mu}$$

$$= \frac{(2\pi)^{1/2} \exp\left(-\frac{1}{2}\sum_i x_i^2\right)}{\exp\left(-\frac{1}{2}\left(1+\sum_i x_i^2\right)\right) \exp\left(\frac{\left(1+\sum x_i\right)^2}{2(n+1)}\right) (2\pi)^{1/2}(n+1)^{-1/2}}$$

$$= (n+1)^{1/2} \exp\left(\frac{1}{2}\right) \exp\left(-\frac{\left(1+\sum x_i\right)^2}{2(n+1)}\right)$$

If $n = 10$ and $\sum x_i = 5$ then $\mathrm{BF} = 1.06$. [Remember: $\frac{1}{\sigma\sqrt{2\pi}} = \int \exp\left(-\frac{1}{\sigma^2}(x-\theta)^2\right) dx$
]

More examples

- Composite vs. composite: $x = (x_1, \ldots, x_{100}) \sim N(\mu, 1000)$ and $\bar{x} = 22$.
  $H_0 : \mu < 20, \quad H_1 : \mu \geq 20$
  If $\mu \sim N(24, 30)$ then $\mu \,\big|\, x \sim N\left(6 + \frac{3}{4} \times 22, 7.5\right) = N(22.5, 7.5)$ Prior odds ratio: $\frac{\pi_0}{\pi_1} = \frac{0.233}{0.767} = 0.303$ Posterior odds ratio: $\frac{P(\mu<20|x)}{P(\mu\geq20|x)} = \frac{0.181}{0.819} = 0.227$

  $$BF = 0.227/0.303 = 0.727$$

  [Classical p-value for this test: $P\left(Z > \frac{22-20}{\sqrt{10}}\right) = .264$] Can you calculate odds ratios in previous example? If $\pi(\mu)$ absolutely continuous then $\pi(\mu = 0) = 0$. Use mixture prior with a Dirac mass at zero.

More examples

Example with Dirac mass: $H_0 : \mu = 0, \quad H_1 : \mu \neq 0$
$\pi(\mu) = \rho\delta_0(\mu) + (1-\rho)\pi_1(\mu)$ with $\rho = P(\mu = 0)$. Then

$$\pi(\mu = 0 \mid x) = \frac{f(x \mid \mu = 0)\rho}{\int f(x \mid \mu)\pi(\mu)d\mu} = \frac{f(x \mid \mu = 0)\rho}{f(x \mid \mu = 0)\rho + (1-\rho)m_1(x)}$$

with

$$m_1(x) = \int_{\mu \in H_1} f(x \mid \mu)\pi_1(\mu)d\mu$$

For $x \sim N\left(\mu, \sigma^2\right)$ and $\pi_1(\mu)$ a $N\left(0, \tau^2\right)$ then

$$\frac{m_1(x)}{f(x \mid 0)} = \frac{\sigma}{\sqrt{\sigma^2 + \tau^2}} \frac{e^{-x^2/2\left(\sigma^2+\tau^2\right)}}{e^{-x^2/2\sigma^2}} = \sqrt{\frac{\sigma^2}{\sigma^2 + \tau^2}} \exp\left\{\frac{\tau^2 x^2}{2\sigma^2\left(\sigma^2 + \tau^2\right)}\right\}$$

More examples

and

$$\pi(\mu = 0 \mid x) = \left[1 + \frac{1-\rho}{\rho}\sqrt{\frac{\sigma^2}{\sigma^2+\tau^2}}\exp\left(\frac{\tau^2 x^2}{2\sigma^2\left(\sigma^2+\tau^2\right)}\right)\right]^{-1}$$

See also variable selection later on ...

The Multinomial Model

- $x = (x_1, \ldots, x_k)$ vector of counts, with x_j = number of observations for the $j-$ th category and $\sum x_j = n$
- $p(x \mid \theta) \propto \prod_{j=1}^{k} \theta_j^{x_j}$ with $\sum_j \theta_j = 1$
- Conjugate prior is Dirichlet, $\theta \sim D(\alpha)$

$$\pi(\theta) \propto \prod_{j=1}^{k} \theta_j^{\alpha_j - 1}$$

$$\pi(\theta \mid x) \propto \prod_{j=1}^{k} \theta_j^{x_j + \alpha_j - 1}$$

that is, $\theta \mid x \sim D(\alpha + x)$

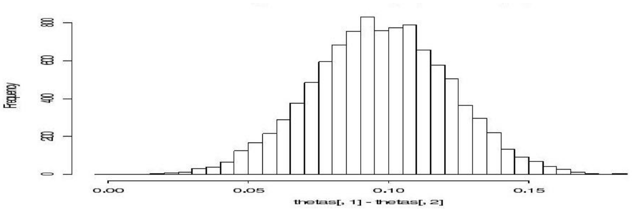
- Uniform prior if $\alpha_j = 1$ for every j. Improper prior if $\alpha_j = 0$ for every j (uniform on $\log \theta_j$) but with proper posterior if there is at least one obs in each of the k categories (i.e., $x_j > 0$ for each j).

Example

Survey of 1447 US voters to find out their preferences in the upcoming presidential election. Data: $x_1 = 727$ support the republican candidate, $x_2 = 583$ the democratic candidate, $x_3 = 173$ have no preference.

$$p(x \mid \theta) \propto \theta_1^{727}\theta_2^{583}\theta_3^{173}, \quad \theta \sim D(1,1,1), \quad \theta \mid x \sim D(728, 584, 174)$$

Histogram of $\theta_1 - \theta_2$ indicates more support for republican candidate




```
thetas <- rdirichlet(10000,c(728,584,174))
hist(thetas[,1]-thetas[,2], nclass=50)
```

Prelude to MCMC methods

Often the posterior distribution of a parameter θ or a function of it, $g(\theta)$, cannot be derived in closed form. Example: Non-conjugate priors. For the normal model $x \sim N\left(\mu, \sigma^2\right)$, the independent prior $\pi\left(\mu, \sigma^2\right) = \pi(\mu)\pi\left(\sigma^2\right)$ with $\mu \sim N\left(\mu_0, \sigma_0^2\right)$ and $\sigma^2 \sim 1/\sigma^2$ (or Inv-Ga) leads to $p\left(\sigma^2 \mid x\right)$ which is not of a familiar form.
Need alternative methods for computing posterior distributions and post summaries.
Available options:

- Analytical methods based on approximations, e.g., large sample normal approximation of the posterior distribution, Laplace approximation methods, numerical integration (will not be covered).
- Simulation methods bases on direct sampling from the posterior (rejection sampling, importance sampling).
- Simulate from a Markov chain whose stationary distribution is the desired posterior distribution (e.g., via Gibbs sampler and Metropolis-Hastings algorithms) and then calculate MC estimates