STAT 425: Introduction to Bayesian Analysis

Marina Vannucci

Rice University, USA
Lecture 7: Multi-parameter models

- The Normal model
- The multinomial
Parameterizations of the Normal Distribution

- Mean and deviation:
  
  \[ f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}, \quad \sigma > 0. \]

- Mean and precision:

  \[ f(x | \mu, \tau) = \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau(x-\mu)^2}{2}}, \quad x \in \mathbb{R}, \quad \tau = \frac{1}{\sigma^2} > 0. \]

- The latter has advantages in numerical computations when \( \sigma \to 0 \) and simplify formulas.
## Summary

<table>
<thead>
<tr>
<th></th>
<th>pdf/pmf</th>
<th>Domain</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bern</strong></td>
<td>$P(x) = p^x (1 - p)^{1-x}$</td>
<td>${0, 1}$</td>
<td>$p$</td>
<td>$p(1 - p)$</td>
</tr>
<tr>
<td><strong>Bin</strong></td>
<td>$P(x) = \binom{N}{x} p^x (1 - p)^{N-x}$</td>
<td>${0, \ldots, N}$</td>
<td>$Np$</td>
<td>$Np(1 - p)$</td>
</tr>
<tr>
<td><strong>Poi</strong></td>
<td>$P(x) = e^{-\lambda} \frac{\lambda^x}{x!}$</td>
<td>$\mathbb{N}$</td>
<td>$\lambda$</td>
<td>$\lambda$</td>
</tr>
<tr>
<td><strong>NB</strong></td>
<td>$P(x) = \binom{r+x-1}{x} p^r (1 - p)^x$</td>
<td>$\mathbb{N}$</td>
<td>$r \frac{1-p}{p}$</td>
<td>$r \frac{1-p}{p^2}$</td>
</tr>
<tr>
<td><strong>M</strong></td>
<td>$P(x_1, \ldots, x_k) = \frac{N!}{\prod_k x_k!} \prod_k p_k^{x_k}$</td>
<td>${0, \ldots, N}^K$</td>
<td>$Np$</td>
<td>$Np_k (1 - p_k)$ $- Np_k p_k'$</td>
</tr>
<tr>
<td><strong>U</strong></td>
<td>$f(x) = \frac{1}{b-a}$</td>
<td>$[a, b]$</td>
<td>$\frac{a+b}{2}$</td>
<td>$\frac{(b-a)^2}{12}$</td>
</tr>
<tr>
<td><strong>Be</strong></td>
<td>$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1 - x)^{b-1}$</td>
<td>$[0, 1]$</td>
<td>$\frac{a}{a+b}$</td>
<td>$\frac{ab}{(a+b)^2(a+b+1)}$</td>
</tr>
<tr>
<td><strong>Ga</strong></td>
<td>$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$</td>
<td>$\mathbb{R}^+$</td>
<td>$\frac{a}{b}$</td>
<td>$\frac{a}{b^2}$</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$</td>
<td>$\mathbb{R}$</td>
<td>$\mu$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td><strong>MN</strong></td>
<td>$f(X) = (2\pi)^{-\frac{p}{2}}</td>
<td>\Sigma</td>
<td>^{-\frac{1}{2}}$ $e^{-\frac{1}{2} (X-\mu)^T \Sigma^{-1} (X-\mu)}$</td>
<td>$\mathbb{R}^p$</td>
</tr>
</tbody>
</table>
## Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>MOM</th>
<th>MLE</th>
<th>UMVUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bern</td>
<td>( p )</td>
<td>( \overline{X} )</td>
<td>( \overline{X} )</td>
<td>( \overline{X} )</td>
</tr>
<tr>
<td>Bin</td>
<td>( p )</td>
<td>( \overline{X} - \frac{n-1}{n} S^2 )</td>
<td>( \frac{\overline{X}}{nN} )</td>
<td>( \frac{\overline{X}}{nN} )</td>
</tr>
<tr>
<td>Poi</td>
<td>( \lambda )</td>
<td>( \overline{X} )</td>
<td>( \overline{X} )</td>
<td>( \overline{X} )</td>
</tr>
<tr>
<td>NB</td>
<td>( r )</td>
<td>( \overline{X} - \frac{n-1}{n} S^2 )</td>
<td>with known ( r )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( p )</td>
<td>( \frac{\hat{r}}{\hat{r} + \overline{X}} )</td>
<td>( \frac{r}{r + \overline{X}} )</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>( a )</td>
<td>( \overline{X} - \sqrt{3} \frac{n-1}{n} S^2 )</td>
<td>( X_{(1)} )</td>
<td>with ( a = 0 )</td>
</tr>
<tr>
<td></td>
<td>( b )</td>
<td>( \overline{X} + \sqrt{3} \frac{n-1}{n} S^2 )</td>
<td>( X_{(n)} )</td>
<td>( \frac{n+1}{n} X_{(n)} )</td>
</tr>
<tr>
<td>Ga</td>
<td>( a )</td>
<td>( \frac{\overline{X}^2}{n \frac{n-1}{n} S^2} )</td>
<td>with known ( a )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( b )</td>
<td>( \frac{\overline{X}}{n \frac{n-1}{n} S^2} )</td>
<td>( \frac{a}{\overline{X}} )</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>( \mu )</td>
<td>( \overline{X} )</td>
<td>( \overline{X} )</td>
<td>( \overline{X} )</td>
</tr>
<tr>
<td></td>
<td>( \sigma^2 )</td>
<td>( \frac{n-1}{n} S^2 )</td>
<td>( \frac{n-1}{n} S^2 )</td>
<td>with known ( \sigma^2 )</td>
</tr>
</tbody>
</table>
Related Distributions

- Normal distribution $X \sim N(\mu, \sigma^2)$:
  - Truncated normal distribution:
    $$f(x|\mu, \sigma^2, a, b) = \frac{f(x|\mu, \sigma^2)}{\Phi \left( \frac{b-\mu}{\sigma} \right) - \Phi \left( \frac{a-\mu}{\sigma} \right)};$$

- Standardized t-distribution:
  $$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}(0, 1), \quad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad s = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

- Standard normal distribution $X \sim N(0, 1)$:
  - Log-normal distribution: $e^{\mu+\sigma X} \sim LN(\mu, \sigma^2)$;
  - Cauchy distribution: $X_1/X_2 \sim Cauchy(0, 1)$;
Bell-shaped Distributions

- Laplace distribution (double exponential distribution):
  \[ f(x|\mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}, \ x \in \mathbb{R}, \ b > 0. \]

- Cauchy distribution:
  \[
  f(x|\mu, \gamma) = \frac{1}{\pi \gamma \left[1 + \left(\frac{x-\mu}{\gamma}\right)^2\right]}, \ x \in \mathbb{R}, \ b > 0.
  \]

- t-distribution:
  \[
  f(x|\nu, \mu, \sigma) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu \pi \sigma \Gamma\left(\frac{\nu}{2}\right)}} \left[1 + \frac{1}{\nu} \left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}}, \ x \in \mathbb{R}, \ \nu > 0, \ \sigma > 0.
  \]

- Logistic distribution:
  \[
  f(x|\mu, s) = \frac{e^{-\frac{x-\mu}{s}}}{s \left(1 + e^{-\frac{x-\mu}{s}}\right)^2}, \ x \in \mathbb{R}, \ s > 0.
  \]
Laplace, Cauchy, Standardized t and logistic
The Gamma distribution - a refresher

- The Gamma distribution is often used to model parameters that can only take positive values.

- In turn, this has been motivated by the fact that the Gamma distribution acts as a conjugate prior in many models

\[ \theta \sim \text{Gamma}(\alpha, \beta) \]

\[ p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta \theta} \quad \alpha, \beta > 0 \]

- \text{Gamma}(1, \beta) \equiv \text{Exp}(\beta)
  (exponential density)

- \text{Gamma}\left(\frac{\nu}{2}, \frac{1}{2}\right) \equiv \chi^2_{\nu}
  (chi-square density)
The Gamma distribution

- The Gamma distribution is often used to model parameters that can only take positive values.
- In turn, this has been motivated by the fact that the Gamma distribution acts as a conjugate prior in many models

\[ \theta \sim \text{Gamma}(\alpha, \beta) \]

\[ p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta \theta} \quad \alpha, \beta > 0 \]

- \( E(\theta) = \frac{\alpha}{\beta} \)
- \( \text{Mode}(\theta) = \frac{\alpha - 1}{\beta}, \alpha > 1 \)
- \( V(\theta) = \frac{\alpha}{\beta^2} \)
Possible models

- **Data likelihood:**

\[
 f(x_1, \ldots, x_n | \mu, \sigma^2) = \prod_{i=1}^{n} f(x_i | \mu, \sigma^2) \\
 = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\
 = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}.
\]

- **Models:**
  - \(\mu\) is unknown, \(\sigma^2\) is known;
  - \(\mu\) is known, \(\sigma^2\) is unknown;
  - Both \(\mu\) and \(\sigma^2\) are unknown:
    - \(\mu\) is dependent on \(\sigma^2\);
    - \(\mu\) and \(\sigma^2\) are independent.
Useful facts for derivations

- Normal component: if \( \pi(\theta) \propto e^{-\frac{1}{2}(a\theta^2 - 2b\theta)} \), then

  \[
  \theta \sim N \left( \frac{b}{a}, \frac{1}{a} \right)
  \]

  and

  \[
  \int \frac{1}{\sqrt{2\pi}} a^{\frac{1}{2}} e^{-\frac{b^2}{2a}} e^{-\frac{1}{2}(a\theta^2 - 2b\theta)} \, d\theta = 1.
  \]

- Gamma component: if \( \pi(\theta) \propto \theta^{a-1} e^{-b\theta} \), then

  \[
  \theta \sim \text{Ga} \left( a, b \right)
  \]

  and

  \[
  \int \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \, d\theta = 1.
  \]
Student component: if \( \pi(\theta) \propto \left( \delta + \frac{(\theta - l)^2}{S} \right)^{-\frac{\delta + 1}{2}} \), then

\[
\theta \sim t_\delta (l, S)
\]

and

\[
\int \frac{1}{\sqrt{\pi S^2}} \frac{\Gamma \left( \frac{\delta + 1}{2} \right)}{\Gamma \left( \frac{\delta}{2} \right)} \delta^\frac{\delta}{2} \left( \delta + \frac{(\theta - l)^2}{S} \right)^{-\frac{\delta + 1}{2}} d\theta = 1.
\]
The Normal Model

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

\[
L(\mu, \sigma^2) \propto \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma^2} (x_i - \mu)^2 \right)
\]

\[
\propto \left( \frac{1}{\sigma^2} \right)^{n/2} \exp \left( -\frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2 \right)
\]

For inference, focus is on \( p(\mu, \sigma^2 | x) = p(\mu | \sigma^2, x)p(\sigma^2 | x) \).

From a Bayesian perspective, it is easier to work with the precision, \( \tau = \frac{1}{\sigma^2} \). The likelihood becomes:

\[
L(\mu, \tau) \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \tau^{1/2} \exp \left( -\frac{1}{2\tau} (x_i - \mu)^2 \right)
\]

\[
\propto \tau^{n/2} \exp \left( -\frac{1}{2\tau} \sum_i (x_i - \mu)^2 \right)
\]
Likelihood factorization:

\[ L(\mu, \tau) \propto \tau^{n/2} \exp \left( -\frac{1}{2} \tau \sum_i (x_i - \mu)^2 \right) \]

\[ \propto \tau^{n/2} \exp \left( -\frac{1}{2} \tau \sum_i [(x_i - \bar{x}) - (\mu - \bar{x})]^2 \right) \]

\[ \propto \tau^{n/2} \exp \left( -\frac{1}{2} \tau \left[ \sum_i (x_i - \bar{x})^2 + n(\mu - \bar{x})^2 \right] \right) \]

\[ \propto \tau^{n/2} \exp \left( -\frac{1}{2} \tau s^2 (n - 1) \right) \exp \left( -\frac{1}{2} \tau n(\mu - \bar{x})^2 \right) \]

\[ \propto \tau^{n/2} \exp \left( -\frac{1}{2} \tau SS \right) \exp \left( -\frac{1}{2} \tau n(\mu - \bar{x})^2 \right) \]

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

Non-informative prior: \( \pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2} \). This arises by considering \( \mu \) and \( \sigma^2 \) 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 \( \mu \) and \( \sigma^2 \) is

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

where

\[
s^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]
The conditional posterior distribution, \( p(\mu|\sigma^2, x) \), is equivalent to deriving the posterior for \( \mu \) when \( \sigma^2 \) is known

\[
\mu|\sigma^2, x \sim \mathcal{N}\left(\bar{x}, \frac{\sigma^2}{n}\right)
\]

The marginal posterior \( p(\sigma^2|x) \), is obtained integrating \( p(\mu, \sigma^2|x) \) over \( \mu \)

[Hint: integral of a Gaussian function \( 2c\sqrt{\pi} = \int 2 \exp\left(-\frac{1}{c^2}(\mu + b)^2\right)d\mu \)]

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

\[
\propto (\sigma^2)^{-\left[(n-1)/2+1\right]} \exp\left\{-\frac{(n - 1)s^2}{2\sigma^2}\right\}
\]

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

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

or, equivalently, \( \tau|x \sim Ga \).
Sampling from the joint posterior distribution

One can simulate a value of \((\mu, \sigma^2)\) from the joint posterior density by

1. simulating \(\sigma^2\) from an inverse-Gamma\((\frac{n-1}{2}, s^2 \frac{n-1}{2})\) distribution [take the inverse of random samples from a Gamma\((\frac{n-1}{2}, s^2 \frac{n-1}{2})\) ]

2. then simulating \(\mu\) from \(\mathcal{N}\left(\bar{x}, \frac{\sigma^2}{n}\right)\) distribution.
Marginal posterior distribution \( p(\mu|x) \) of \( \mu \)

As \( \mu \) is typically the parameter of interest (\( \sigma^2 \) nuisance parameter) it is useful to calculate its marginal posterior distribution

[Hint: integral of a Gamma function \( \frac{\Gamma(a)2^a}{ba} = \int_0^\infty z^{a-1} \exp(-\frac{zb}{2})dz \)]

\[
p(\mu|x) = \int_0^\infty p(\mu, \sigma^2|x)\ d\sigma^2
\]

\[
\propto \int_0^\infty (\sigma^2)^{-\left(\frac{n}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}\left[(n - 1)s^2 + n(\bar{x} - \mu)^2\right]\right\} d\sigma^2
\]

\[
= A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z)dz,
\]

with \( A = (n - 1)s^2 + n(\bar{x} - \mu)^2 \), \( z = \frac{A}{2\sigma^2} \)

\[
\propto A^{-n/2} = \left[1 + \frac{1}{n - 1} \left(\frac{\mu - \bar{x}}{s/\sqrt{n}}\right)^2\right]^{-\left[(n-1)+1\right]/2}
\]

that is, \( \mu|x \sim t(n - 1, \bar{x}, s^2/n) \), or \( \frac{\mu - \bar{x}}{s/\sqrt{n}} |x \sim t_{n-1} \) with \( t_{n-1} \) the standard \( t \)-distribution with \( n - 1 \) degrees of freedom.
Conjugate Prior Model

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

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

which corresponds to the joint prior density

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

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

we call this a Normal-Inverse-Gamma prior,

$$(\mu, \sigma^2) \sim NIG(\mu_0, \tau_0, \nu_0/2, SS_0/2)$$
Joint Posterior $p(\mu, \sigma^2 | y)$

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

with

\[
\mu | \sigma^2, x \sim N(\mu_n, \sigma^2 / \tau_n), \quad \mu_n = \frac{\mu_0 \tau_0}{\sigma^2} + \bar{x} \frac{n}{\sigma^2} = \frac{\tau_0 \mu_0 + n \bar{x}}{\tau_n}, \quad \tau_n = \tau_0 + n
\]

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

Thus, $\mu, \sigma^2 | y \sim \text{Normal-Inverse Gamma}(\mu_n, \tau_n; \nu_n/2, SS_n^2/2)$. 
Also $\mu | x \sim t_{\nu_n}(\mu_n, \sigma_n^2/\tau_n)$, $\sigma_n^2 = SS_n^2/\nu_n$

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

Comments:

- $\mu_n$ expected value for $\mu$ after seeing the data
  \[ \mu_n = \frac{n}{\tau_n} \bar{x} + \frac{\tau_0}{\tau_n} \mu_0, \text{ weighted average} \]

- $\tau_n$ precision for estimating $\mu$ after $n$ observations.

- $\nu_n$ degrees of freedom
  \[ [\tau \sim Ga(\alpha/2, \beta/2) \rightarrow \beta \tau \sim \chi^2_\alpha, \text{ with } \alpha \text{ degrees of freedom}] \]

- $SS_n$ posterior variation as prior variation+observed variation+variation between prior mean and sample mean.

- Limiting case $\tau_0 \rightarrow 0, \nu_0 \rightarrow -1$ (and $SS \rightarrow 0$) then
  \[ \mu | x \sim t_{n-1}(\bar{x}, s^2/n) \text{ (same as improper prior!)} \]
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 $\log(\text{ratios}) = \text{difference in logs}$). Ratios make more sense given the goals: how much longer can a person be exposed to the sun relative to their baseline.

Model: $Y = \log(\text{TRT}) - \log(\text{CONTROL}) \sim N(\mu, \tau)$. Then $E(\log(\text{TRT}/\text{CONTROL})) = \mu = \log(\text{SPF})$. Interested in $\exp(\mu) = \text{SPF}$.

Summary statistics: $\bar{y} = 1.998, s^2 = 0.525, n = 13$

[make boxplots and Q-Q normal plots to check on normality]
Model formulation: \( Y = \log(\text{TRT}) - \log(\text{CONTROL}) \sim N(\mu, \sigma^2), \) 
\( n = 13, \bar{y} = 1.998, SS = 0.525. \)

Question: \( \pi(\mu|y_1, \ldots, y_n) = ? \)

Bayesian model:
- Data likelihood: \( f(y_1, \ldots, y_n|\mu, \sigma^2) = \prod_{i=1}^{n} N(y_i; \mu, \sigma^2); \)
- Non-informative Prior: \( (\mu, \sigma^2) \sim 1/\sigma^2; \)
- Posterior: \( (\mu, \sigma^2|y_1, \ldots, y_n) \sim N(\bar{y}, \sigma^2/n) IG\left(\frac{n-1}{2}, \frac{s^2}{2n-1}\right) \)
- Posterior: \( \mu|y_1, \ldots, y_n \sim t_{n-1}(\bar{y}, \frac{s^2}{n}); \)
- Prediction: \( y_f|y_1, \ldots, y_n \sim t_{n-1}(\bar{y}, s^2(n - 1)/n). \)

Coding in R: rgamma(), rnorm() and rt().
With non-informative prior.

Posterior: \((\mu, \sigma^2 | y_1, \ldots, y_n) \sim N(\bar{y}, \sigma^2 / n) IG\left(\frac{n-1}{2}, s^2 \frac{n-1}{2}\right)\)

Posterior: \(\mu | y_1, \ldots, y_n \sim t_{n-1}(\bar{y}, \frac{s^2}{n})\)

Define: \(vn = (n - 1) = 12, SSn = s^2(n - 1) = 0.525, mn = 1.998\)

Sampling from posterior:

Draw \(\tau | Y\)

\(\tau = \text{rgamma}(10000, \ vn/2, \ \text{rate}=SSn/2)\)

Draw \(\mu | \tau, Y\)

\(\mu = \text{rnorm}(10000, \ mn, 1/\sqrt{\phi \times n})\)

or draw \(\mu | Y\) directly

\(\mu = \text{rt}(10000, vn) \times \sqrt{SSn/(n \times vn)} + mn\)
Model formulation: \( Y = \log(\text{TRT}) - \log(\text{CONTROL}) \sim N(\mu, \sigma^2), \) 
\( n = 13, \bar{y} = 1.998, SS = 0.525. \)

Question: \( \pi(\mu|y_1, \ldots, y_n) =? \)

Bayesian model:
- Data likelihood: \( f(y_1, \ldots, y_n|\mu, \sigma^2) = \prod_{i=1}^{n} N(y_i; \mu, \sigma^2); \)
- Conjugate Prior: \( \mu|\sigma^2 \sim N(\mu_0, \frac{\sigma^2}{\tau_0}), \sigma^2 \sim IG(\frac{\nu_0}{2}, \frac{SS_0}{2}); \)
- Posterior: \( (\mu, \sigma^2|y_1, \ldots, y_n) \sim NIG(\mu_n, \tau_n; \nu_n/2, SS_n^2) \)
- Posterior: \( \mu|y_1, \ldots, y_n \sim t_{\nu_n}(\mu_n, \frac{SS_n}{\tau_n \nu_n}); \)
- Prediction: \( y_f|y_1, \ldots, y_n \sim t_{\nu_n}(\mu_n, \frac{SS_n}{\nu_n \frac{\tau_n+1}{\tau_n}}). \)

Coding in R: rgamma(), rnorm() and rt().
Expert opinions on $\mu$:
· Best guess on median SPF is 16
· $P(\mu > 64) = 0.01$
· information in prior is worth 25 observations

Possible subjective prior: $\mu_0 = \log(16)$, $\tau_0 = 25$, $\nu_0 = \tau_0 - 1$
$P(\mu < \log(64)) = .99$ implies $SS_0 = 185.7$

Posterior hyperpar: $\tau_n = 38$, $\mu_n = 2.508$, $\nu_u = 37$, $SS_n = 197.134$

Sampling from posterior:
Draw $\tau|Y$

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

Draw $\mu|\tau, Y$

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

or draw $\mu|Y$ directly

mu = rt(10000, vn) * sqrt(SSn/(tn*vn)) + mn
Transform to $\exp(\mu)$. Find 95% C.I. of 4.54 to 23.758
Predictive Distribution of future $z$

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

  $$p(z|x) = \int \int p(z|\mu, \sigma^2, x)p(\mu, \sigma^2|x) \, d\mu d\sigma^2$$

  [Use assumption that $z$ is independent of $x$ given $\mu$ and $\sigma^2$, then integrate $\mu$ using the normal integral, then integrate $\sigma^2$ using the Gamma integral]

- Reference prior: $z|x \sim t_{n-1}\left(\bar{x}, s^2(n + 1)/n\right)$

- Conjugate prior: $z|x \sim t_{\nu_n}\left(\mu_n, \sigma^2_n(\tau_n + 1)/\tau_n\right), \sigma^2_n = SS^2_n/\nu_n$

  [Can use the normal “trick” to integrate $\mu$: If $z \sim N(\mu, \sigma^2)$ and $\mu \sim N(\mu_0, \sigma^2/\tau_0)$ then $y = \frac{z-\mu}{\sigma} \sim N(0, 1)$, that is $z = \sigma y + \mu$ and therefore $z|\sigma^2 \sim N(\mu_0, \sigma^2(1 + \frac{1}{\tau_0}))$ 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,

\[ p(z) = \int \int p(z|\mu, \sigma^2) \pi(\mu, \sigma^2) d\mu d\sigma^2 \rightarrow z \sim t_{\nu_0}(\mu_0, \frac{SS_0}{\nu_0}(1 + \frac{1}{\tau_0})) \]

[as above]

\[ [\int \int N(\mu, \sigma^2)N(\mu_0, \sigma^2/\tau_0)IG(\nu/2, SS/2)d\mu d\sigma^2 = t_{\nu}(\mu_0, \frac{SS}{\nu}(1 + \frac{1}{\tau_0}))] \]

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

\[ Y = r t(10000, 24) \ast \sqrt{\left( 1 + \frac{1}{25} \right) \ast 187.5 / 24} + \log(16) \]

quantile(exp(Y))

0% 25% 50% 75% 100%
4.57e-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

\[ \pi(\mu, \sigma^2) = \pi(\mu)\pi(\sigma^2) \]

\[ \mu \sim N(\mu_0, \sigma_0^2), \quad \sigma^2 \sim IG\left(\frac{\nu_0}{2}, \frac{SS_0}{2}\right) \]

then \[ \mu|\sigma^2, x \sim N(\mu_n, \tau_n^2), \quad \mu_n = \frac{\mu_0}{\sigma_0^2} + \frac{\bar{x} n}{\sigma^2}, \quad \tau_n^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \]

\[ \sigma^2|x \sim \text{not in closed form} \]

We will solve this with MCMC methods!
Summary of Conjugate Priors for the Normal Model

- Conjugate priors for normal data with unknown precision are

\[ \tau \sim \text{Gamma} \left( \frac{a}{2}, \frac{b}{2} \right) \quad \mu | \tau \sim N \left( \mu_0, \frac{1}{\tau_0 \tau} \right) \]

Here $a$, $b$, $\mu_0$, and $\tau_0$ are known hyper-parameters chosen to characterize the prior information.

- The problem with using this prior in practical data analysis is the difficulty of specifying a distribution for $\mu$ that is conditional on $\tau$ (which is also unknown).
Summary of Independence prior

- Here we assume that information about $\mu$ can be elicited independently of information on $\tau$ or $\sigma$, so

  \[ p(\mu, \tau) = p(\mu) p(\tau) \]

- This makes elicitation relatively easy. Although the primary goal is to get a prior that reasonably captures the expert’s information, independence priors work generally well.

- Usually, one considers Gamma priors for $\tau$, since they are conjugate. But there’s really no need, as long as the prior is defined on the positive real line.
Proper (semi-conjugate) Reference Priors

More recently priors such as

\[ \mu \sim N(0, b) \]
\[ \tau \sim \text{Gamma}(c, c) \]

have been used as proper reference priors.

In this case, \( b \) and \( c \) are chosen so that the prior precision for \( \mu, 1/b \), and both hyperparameters \( c \) in the Gamma distribution are near zero.

Such priors are seen as approximation of the \( p(\mu, \tau) \propto 1/\tau \) improper default prior.

Common choices are \( b = 10^6 \) and \( c = 0.001 \).
Back to the example

- We need to identify a prior distribution that gives information/no-information about the unknown parameters $\mu$ and $\tau = 1/\sigma^2$.

- $\mu \sim N(0, 10^6)$ as proper non-informative prior.

- Expert opinion that $\mu$ should be centered at 16. Then, $\mu \sim N(16, 10^6)$ as diffuse prior.

- Expert 95% certain that the mean SPF should be $\mu$ should be between 10 and 75, that is, $Pr(10 < \mu < 75) = 0.95$. Then

  $\mu \sim N(10, 0.0163)$
Back to the example

- We have no good information on $\sigma^2$, the variance of an observation.
- So we can specify a reference (vague) prior on $\tau$, which is independent of $\mu$:

$$\tau \sim Gamma(0.001, 0.001)$$
The Multinomial Model

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

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

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

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

- 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.

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

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

```r
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 $\theta$ or a function of it, $g(\theta)$, cannot be derived in closed form.

Example: Non-conjugate priors. For the normal model $x \sim N(\mu, \sigma^2)$, the independent prior $\pi(\mu, \sigma^2) = \pi(\mu)\pi(\sigma^2)$ with $\mu \sim N(\mu_0, \sigma_0^2)$ and $\sigma^2 \sim 1/\sigma^2$ (or Inv-Ga) leads to $p(\sigma^2 | x)$ 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.