Law of Total Probability

If $A_1, A_2, \dots$ form a *partition* of Ω and $B \in \Omega$, then $\mathrm{P}(B) = \sum_{n=1}^\infty \mathrm{P}(B \cup A_n) = \sum_{n=1}^\infty \mathrm{P}(B \mid A_n)\,\mathrm{P}(A_n)$.

Bayes' Rule

- For $A, B \in \mathcal{F}$, $\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)P(A)}{\mathrm{P}(B)} = \frac{\mathrm{P}(B \mid A)P(A)}{\mathrm{P}(B \mid A)\,\mathrm{P}(A)+\mathrm{P}(B \mid A^c)\,\mathrm{P}(A^c)}$.
- For a partition $\{A_1, A_2, \dots\}$ and $B \in \mathcal{F}$, $\mathrm{P}(A_i \mid B) = \frac{\mathrm{P}(B \mid A_i)P(A_i)}{\sum_{i=1}^\infty \mathrm{P}(B \mid A_i)\,\mathrm{P}(A_i)}$.
- For a pdf or pmf $p(\cdot)$, $f(\theta \mid y) = \frac{f(y \mid \theta)f(\theta)}{f(y)} = \frac{f(y \mid \theta)f(\theta)}{\int f(y \mid \theta)f(\theta)d\theta}$.

Bayesian Learning

If $Y_i \mid \theta \overset{\text{iid}}{\sim} p(y \mid \theta)$, $p(\theta \mid y_1, \dots, y_i) \propto p(y_i \mid \theta)p(\theta \mid y_1, \dots, y_{i-1}$.

Prediction

$$p(y^* \mid y) = \int p(y^*, \theta \mid y)d\theta = \int p(y^* \mid y, \theta)p(\theta \mid y)d\theta$$
$$= \int \underbrace{p(y^* \mid \theta)}_{\text{model}}\underbrace{p(\theta \mid y)}_{\text{posterior}}d\theta \quad (y^* \perp y).$$

Point Estimation

Loss $L(\theta, \hat\theta)$	$\hat\theta_{\text{Bayes}} = \operatorname{argmin}_{\hat\theta} \mathrm{E}\left[L(\theta, \hat\theta) \mid y\right]$
$(\theta - \hat\theta)^2$	$\mathrm{E}(\theta \mid y)$
$\lvert\theta - \hat\theta\rvert$	$\int_{\hat\theta_{\text{Bayes}}}^\infty p(\theta \mid y)d\theta = 1/2$
$\lim_{\varepsilon \to 0} I\left(\lvert\theta - \hat\theta\rvert < \varepsilon\right)$	$\operatorname{argmax}_\theta p(\theta \mid y)$

Interval Estimation

A $100(1-\alpha)\%$ credible interval is any (L, U) such that $1 - \alpha = \int_L^U p(\theta \mid y)d\theta$.

- Equal-tailed: $\alpha/2 = \int_{-\infty}^L p(\theta \mid y)d\theta = \int_U^\infty p(\theta \mid y)d\theta$,
- One-sided: $L = -\infty$ or $U = \infty$,
- HPD: $p(L \mid y) = p(U \mid y)$ for a unimodal posterior (shortest).

Conjugate Priors

A prior $p(\theta)$ is conjugate if for $p(\theta) \in \mathcal{P}$ and $p(y \mid \theta) \in \mathcal{F}$, then $p(\theta \mid y) \in \mathcal{P}$. A prior is natural conjugate if the prior and likelihood have the same functional form.

- E.g., discrete prior $\mathrm{P}(\theta = \theta_i) = p_i$ where $\sum_{i=1}^I p_i = 1$, then $\mathrm{P}(\theta = \theta_i \mid y) = \frac{p_i p(y \mid \theta_i)}{\sum_{i'=1}^I p_{i'} p(y \mid \theta_{i'})}$, which is also discrete.
- E.g., mixture of conjugate priors $\theta \sim \sum_{i=1}^I p_i p_i(\theta)$ where $\sum_{i=1}^I p_i = 1$, then $\theta \mid y \sim \sum_{i=1}^I p_i' p_i(\theta \mid y)$ where $p_i' \propto p_i \int p(y \mid \theta)p_i(\theta)d\theta$ and $p_i(\theta \mid y) \propto p(y \mid \theta)p_i(\theta)$.

Jeffreys Prior

Noninformative priors may not be so for transformations! Note that $\log\left(\frac{\theta}{1-\theta}\right) \sim \mathrm{Beta}(1,1) \overset{\text{d}}{=} \mathrm{Unif}(0,1)$ implies $\theta \sim \mathrm{Beta}(0,0)$.
This motivates Jeffreys prior, which is invariant to parameter transformation, defined as $p(\theta) \propto \sqrt{\det\left(\mathcal{I}(\theta)\right)}$, where $\mathcal{I}(\theta) = -\mathrm{E}_{y \mid \theta}\left[\frac{\partial^2}{\partial\theta^2}\log p(y \mid \theta)\right]$.

Normal Model (known variance)

For $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, s^2)$,

- Jefferys prior (improper) $\mu \sim \mathrm{Unif}(-\infty, \infty)$ gives $\mu \mid y \sim \mathcal{N}(\bar y, s^2/n)$.
- Natural conjugate prior $\mu \sim \mathcal{N}(m, C)$ gives $\mu \mid y \sim \mathcal{N}(m', C')$ where $C' = \left[\frac{1}{C} + \frac{n}{s^2}\right]^{-1}$ and $m' = C'\left[\frac{1}{C}m + \frac{n}{s^2}\bar y\right]$.
- Note $\mathcal{N}(m, C) \to \mathrm{Unif}(-\infty, \infty)$ as $C \to \infty$.

Normal Model (known mean)

For $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(m, \sigma^2)$,

- Jefferys prior (improper) $p(\sigma^2) \propto 1/\sigma^2 \overset{\text{d}}{=} \mathrm{InvGam}(0,0)$ gives $\sigma^2 \mid y \sim \mathrm{InvGam}\left(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^n (y_i - m)^2\right)$.
- Natural conjugate prior $\sigma^2 \sim \mathrm{InvGam}(a, b)$ gives $\sigma^2 \mid y \sim \mathrm{InvGam}\left(a + \frac{n}{2}, b + \frac{1}{2}\sum_{i=1}^n (y_i - m)^2\right)$.

Scaled Inv-χ^2

- $\sigma^2 \sim \mathrm{InvGam}(a, b) \overset{\text{d}}{=} \text{Inv-}\chi^2(v, s^2)$ where $a = v/2$ and $b = vs^2/2$ (equivalently, $v = 2a$ and $s^2 = b/a$).
- Mean: $\frac{vs^2}{v-2}$ for $v > 2$; Mode: $\frac{vs^2}{v+2}$; Variance: $\frac{2v^2(s^2)^2}{(v-2)^2(v-4)}$ for $v > 4$.

Location-scale t

- If $T \sim t_v$, then $X = m + sT \sim t_v(m, s^2)$.
- $f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi s}}\left(1 + \frac{1}{v}\left[\frac{x-m}{s}\right]^2\right)^{-\frac{v+1}{2}}, \quad x \in \mathbb{R}.$
- $t_v(m, s^2) \to \mathcal{N}(m, s^2)$ as $v \to \infty$.

Normal Inv-χ^2

- If $\mu \mid \sigma^2 \sim \mathcal{N}(m, \sigma^2/k)$ and $\sigma^2 \sim \text{Inv-}\chi^2(v, s^2)$, then $p(\mu, \sigma^2) \propto (\sigma^2)^{-\frac{v+3}{2}}\exp\left[-\frac{1}{2\sigma^2}\left(k(\mu - m)^2 + vs^2\right)\right]$.
- Marginally, $\mu \sim t_v(m, s^2/k)$.

Normal model

For $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$,

- Jefferys prior $p(\mu, \sigma^2) \propto (1/\sigma^2)^{3/2}$ (not often used).
- Reference prior (improper) $p(\mu, \sigma^2) \propto 1/\sigma^2 \overset{\text{d}}{=} \mathrm{InvGam}(0,0)$ gives $\mu \mid \sigma^2, y \sim \mathcal{N}(\bar y, \sigma^2/n)$ and $\sigma^2 \mid y \sim \text{Inv-}\chi^2\,(n-1, s^2)$. Marginally, $\mu \mid y \sim t_{n-1}(\bar y, s^2/n)$.
- Predictive $\int\int p(y^* \mid \mu, \sigma^2 p(\mu \mid \sigma^2 y)p(\sigma^2 \mid y)d\mu d\sigma^2$. Trick is to write $y^* = \mu + \varepsilon$ where $\mu \mid \sigma^2, y \sim \mathcal{N}(\bar y, \sigma^2/n)$ independent of $\varepsilon \mid \sigma^2, y \sim \mathcal{N}(0, \sigma^2)$. Then $y^* \mid \sigma^2, y \sim \mathcal{N}(\bar y, \sigma^2(1 + 1/n))$ with $\sigma^2 \mid y \sim \text{Inv-}\chi^2(n-1, s^2)$ so that $y^* \mid y \sim t_{n-1}(\bar y, s^2(1 + 1/n))$.
- Natural conjugate prior $\mu \mid \sigma^2 \sim \mathcal{N}(m, \sigma^2/k)$ and $\sigma^2 \sim \text{Inv-}\chi^2(v, s^2)$ gives $\mu \mid \sigma^2, y \sim \mathcal{N}(m', \sigma^2/k')$ and $\sigma^2 \mid y \sim \text{Inv-}\chi^2(v', (s')^2)$ where $k' = k + n$, $v' = v + n$, $m' = \frac{km + n\bar y}{k'}$, and $v'(s')^2 = vs^2 + (n-1)s^2 + \frac{kn}{k'}(\bar y - m)^2$. Marginally, $\mu \mid y \sim t_{v'}(m', (s')^2/k')$.

Multinomial

- $Y = (Y_1, \dots, Y_n) \sim \mathrm{Multi}(n, \pi)$ where $\sum_{i=1}^n \pi_i = 1$.
- Multivariate pmf $f(y_1, \dots, y_n) = n!\prod_{i=1}^n \frac{\pi_i^{y_i}}{y_i!}$.
- $\mathrm{E}(Y_i) = n\pi_i$; $\mathrm{Var}(Y_i) = n\pi_i(1 - \pi_i)$; $\mathrm{Cov}(Y_i, Y_{i'}) = -\pi_i\pi_{i'}$.
- Marginally, $Y_i \sim \mathrm{Bin}(n, \pi_i)$.

Dirichlet

- $\pi = (\pi_1, \dots, \pi_K) \sim \mathrm{Dir}(a)$ where $a_k > 0\ \forall k = 1, \dots, K$.
- Multivariate pdf $f(\pi_1, \dots, \pi_K) = \frac{1}{\mathrm{Beta}(a)}\prod_{k=1}^K \pi_k^{a_k - 1}$ where $\sum_{k=1}^J \pi_k = 1$ and $\mathrm{Beta}(a) = \frac{\prod_{k=1}^K \Gamma(a_k)}{\Gamma(a_0)}$ for $a_0 = \sum_{k=1}^K a_k$.
- $\mathrm{E}(\pi_k) = a_k/a_0$; $\mathrm{Var}(\pi_k) = \frac{a_k(a_0 - a_k)}{a_0^2(a_0 + 1)}$; $\mathrm{Cov}(\pi_k, \pi_{k'}) = \frac{-a_k a_{k'}}{a_0^2(a_0 + 1)}$.
- Marginally, $\pi_k \sim \mathrm{Beta}(a_k, a_0 - a_k)$.

Multinomial model

- Conjugate prior $\pi \sim \mathrm{Dir}(a)$ gives $p(\pi \mid y) \sim \mathrm{Dir}(a + y)$.
- Jeffreys prior uses $a_k = 1/2\ \forall k$, possibly putting too much mass on rare categories.

MVN

- $p(y) = (2\pi)^{-k/2}\,|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(y - \mu)'\Sigma^{-1}(y - \mu)\right]$.
- $\mathrm{E}(Y_k) = \mu_k$; $\mathrm{Var}(Y_k) = \Sigma_{kk}$; $\mathrm{Cov}(Y_k, Y_{k'}) = \Sigma_{kk'}$.
- Marginally, $Y_k \sim \mathcal{N}(\mu_k, \Sigma_{kk})$.
- If $\begin{pmatrix}Y_1\\ Y_2\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right)$, then $Y_1 \mid Y_2 = y_2 \sim \mathcal{N}(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.
- $\Sigma_{kk'} = 0 \implies Y_k \perp Y_{k'}$ and $\left[\Sigma^{-1}\right]_{kk'} = 0 \implies Y_k \perp Y_{k'} \mid Y_i\ \forall i \neq k, k'$.

MVN model (unknown mean)

If $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, S)$, then

- Default prior $\mu \sim \mathrm{Unif}(-\infty, \infty)$ gives (when $n \geq K$) $\mu \mid y \sim \mathcal{N}(\bar y, S/n)$ where $\bar y$ has elements $\bar y_k = \frac{1}{n}\sum_{i=1}^n y_{ik}$.
- Conjugate prior $\mu \sim \mathcal{N}(m, C)$ gives $\mu \mid y \sim \mathcal{N}(m', C')$ where $C' = [C^{-1} + nS^{-1}]^{-1}$ and $m' = C'\,[C^{-1}m + nS^{-1}\bar y]$.

Inverse Wishart

- $\Sigma_{K \times K} \sim \mathrm{InvWish}(v, W^{-1})$ for $v > K - 1$ and positive definite W if $p(W) \propto \frac{1}{2}\,|W|^{v-K-1}\exp\left[-\frac{1}{2}\operatorname{tr}(W\Sigma^{-1})\right]$.
- $\mathrm{E}(\Sigma) = (v - K - 1)^{-1}W$
- Marginally, $\sigma_k^2 = \Sigma_{kk} \sim \chi^2(v, W_{kk})$.
- $W \sim \mathrm{Wish}(v, S) \implies W^{-1} \sim \mathrm{InvWish}(v, S^{-1})$.
- Problematic as a prior: biases small variances upward.

Normal Inverse Wishart

- $\mu \mid \Sigma \sim \mathcal{N}(m, \frac{1}{c}\Sigma)$ and $\Sigma \sim \text{InvWish}(v, W^{-1})$.
- Marginally, $\mu \sim t_{v-K+1}(m, \frac{1}{c(v-K+1)}W)$.

MVN model

If $Y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \Sigma)$,

- Conjugate prior $\mu \mid \Sigma \sim \mathcal{N}(m, \frac{1}{c}\Sigma)$ and $\Sigma \sim \text{InvWish}(v, W^{-1})$, that is, $p(\mu, \Sigma) \propto |\Sigma|^{-(\frac{v+K}{2}+1)} \exp\left[-\frac{1}{2}\operatorname{tr}(W\Sigma^{-1}) - \frac{c}{2}(\mu - m)'\Sigma^{-1}(\mu - m)\right]$, gives a Normal Inverse Wishart posterior with parameters $c' = c_n$, $v' = v + n$, $m' = \frac{k}{k+n}m + \frac{n}{k+n}\bar{y}$, and $W' = W + S + \frac{kn}{k+n}(\bar{y} - m)(\bar{y} - m)'$ where $S = \sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})'$.
- Default priors: $\Sigma \sim \text{InvWish}(k+1, I)$ marginally puts $\text{Unif}(-1, 1)$ on each correlation; $p(\mu, \Sigma) \propto |\Sigma|^{-(K+1)/2}$ is a Normal Inverse Wishart where $c \to 0, v \to -1, |W| \to 0$ gives $\mu \mid \Sigma, y \sim \mathcal{N}(\bar{y}, \Sigma/n)$ and $\Sigma \mid y \sim \text{InvWish}(n-1, S^{-1})$.

Normal Approximation to Posterior

For unimodal and roughly symmetric $p(\theta \mid y)$, a Taylors expansion about the posterior mode $\hat{\theta}$ gives
$\log p(\theta \mid y) \approx \log p(\hat{\theta} \mid y) - \frac{1}{2}(\theta - \hat{\theta})' \left[-\frac{d^2}{d\theta^2}\log p(\theta \mid y)\right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})$,
that is $p(\theta \mid y) \overset{d}{\approx} \mathcal{N}(\hat{\theta}, I(\hat{\theta})^{-1})$ where
$I(\hat{\theta}) = -\frac{d^2}{d\theta^2}\log p(\theta)\Big|_{\theta=\hat{\theta}} - \frac{d^2}{d\theta^2}\log p(\theta \mid y)\Big|_{\theta=\hat{\theta}}$.

Convergence of Posterior

For $Y_i \overset{\text{iid}}{\sim} p(y \mid \theta_0)$,

- if Θ is discrete and $\text{P}(\theta = \theta_0) > 0$, then $\text{P}(\theta = \theta_0 \mid y) \to 1$ as $n \to \infty$.
- if Θ is continuous and A is a neighborhood around θ_0 such that $\text{P}(\theta \in A) > 0$, then $\text{P}(\theta = A \mid y) \to 1$ as $n \to \infty$.

Consistency of Posterior Estimates

Use fact that $\hat{\theta}_{MLE} \overset{p}{\to} \theta_0$, meaning $\forall \varepsilon > 0$, $\lim_{n\to\infty} \text{P}\left(\left|\hat{\theta} - \theta_0\right| < \varepsilon\right) = 1$.

What can go wrong?

- *Unidentified parameters, parameters grows with n, aliasing, unbounded or heavy-tailed likelihoods.*
- Improper posterior.
- Prior excludes point of convergence.
- Convergence to edge of parameter space (prior).
- Wrong sampling distribution: posterior $p(\theta \mid y)$ will to converge to the point that minimizes Kullback-Leibler divergence from true $f(y)$ where $KL\left[f(y) : p(y \mid \theta)\right] = \text{E}_Y\left[\log\left(\frac{f(y)}{p(y \mid \theta)}\right)\right] = \int \log\left(\frac{f(y)}{p(y \mid \theta)}\right) f(y)dy$.

Hierarchical Models

$$y_i \overset{\text{ind}}{\sim} p(y \mid \theta_i), \quad \theta_i \overset{\text{iid}}{\sim} p(\theta \mid \phi), \quad \phi \sim p(\phi)$$
$$\implies p(\theta, \phi \mid y) \propto p(y \mid \theta, \phi)p(\theta, \phi) \propto p(y \mid \theta, \phi)p(\theta \mid \phi)p(\phi).$$

Can decompose joint posterior as $p(\theta, \phi \mid y) = p(\theta \mid \phi, y)p(\phi \mid y)$, where $p(\theta \mid \phi, y) \propto \prod_{i=1}^n p(\theta_i \mid \phi, y_i)$; $p(\theta \mid y) \propto p(y \mid \phi)p(\theta)$; $p(y \mid \phi) = \prod_{i=1}^n p(y_i \mid \phi)$.
Conditional independencies: $y_i \perp y_j \mid \theta$; $y_i \perp y_j \mid \phi$; $\theta_i \perp \theta_j \mid \phi$; $\theta_i \perp \theta_j \mid \phi, y$; $y \perp \phi \mid \theta$.

Exchangeability

$Y_1, \ldots, Y_n$ are exchangeable if $p(y_1, \ldots, y_n)$ is invariant to permutation of indices (infinitely exchangeable if it holds for any given n).

de Finetti's Theorem

A sequence of r.v.s are infinitely exchangeable iff $\forall n$, $p(y_1, \ldots, y_n) = \int \prod_{i=1}^n p(y_i \mid \theta)P(d\theta)$, for some measure P on θ. If the dsn on θ has a density, can replace $P(d\theta)$ with $p(\theta)d\theta$.

Beta-Binomial Hierarchical Model

- $y_i \overset{\text{ind}}{\sim} \text{Bin}(n_i, \theta_i)$, $\theta_i \overset{\text{iid}}{\sim} \text{Beta}(\alpha, \beta)$, $\alpha, \beta \sim p(\alpha, \beta)$.
- Conditional posterior: $p(\theta \mid \alpha, \beta, y) = \prod_{i=1}^n \text{Beta}(\alpha + y_i,\ \beta + n_i - y_i)$.
- Marginal posterior: $p(\alpha, \beta \mid y) \propto p(y \mid \alpha, \beta)p(\alpha, \beta)$ where $y_i \mid \alpha\beta \overset{\text{ind}}{\sim} \text{Beta-Binomial}(n_i, \alpha, \beta)$, that is $p(y \mid \alpha, \beta) = \prod_{i=1}^n \frac{B(\alpha + y_i, \beta + n_i - y_i)}{B(\alpha, \beta)}$.
- Interpretation: α prior successes, β prior failures. Can parameterize as $\mu = \frac{\alpha}{\alpha+\beta}$ prior expectation, $\eta = \alpha + \beta$ prior sample size. Transform to real line $\log(\mu/[1-\mu]) = \log(\alpha/\beta)$ and $\log \eta$ so we can use $\text{Beta}(20, 30)$ and $\text{Log}\mathcal{N}(0, 3^2$ priors for example (assume they are independent).
- Priors for α and β: warning! $p(\log(\alpha/\beta), \log(\alpha + \beta)) \propto 1$ gives an improper posterior! And $\sim \text{Unif}([-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}])$ is very informative. Can use $\propto \alpha\beta(\alpha + \beta)^{-5/2}$ which leads to a proper posterior and is equivalent to $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$.

Normal Hierarchical Model

- $y_{ij} \overset{\text{ind}}{\sim} \mathcal{N}(\theta_i, \sigma^2)$, $\theta_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \tau^2)$, $\quad \mu, \tau^2 \sim p(\mu, \tau)$.
- For known $\sigma^2 = s^2$, assume $p(\mu, \tau) \propto p(\mu \mid \tau)p(\tau) \propto p(\tau)$, that is, assume improper uniform on μ. Then the posterior is
$$p(\tau \mid y) \propto p(\tau)V_\mu^{1/2} \prod_{i=1}^{I} (s_i^2 + \tau^2)^{-1/2} \exp\left[-\frac{(\bar{y}_{i\cdot} - \hat{\mu})^2}{2(s_i^2 - \tau^2)}\right],$$
$$\mu \mid \tau, y \sim \mathcal{N}(\hat{\mu}, V_\mu), \qquad \theta_i \mid \mu, \tau, y \sim \mathcal{N}(\hat{\theta}_i, V_i),$$
where $V_\mu^{-1} = \sum_{i=1}^I \frac{1}{s_i^2 + \tau^2}$, $V_i^{-1} = 1/s_i^2 + 1/\tau^2$,
$\hat{\mu} = V_\mu\left(\sum_{i=1}^I \frac{\bar{y}_{\cdot i}}{s_i^2 + \tau^2}\right)$, and $\hat{\theta}_i = V_i\left(\frac{\bar{y}_{i\cdot}}{s_i^2} + \frac{\mu}{\tau^2}\right)$.
- Other priors: $\theta_i \sim t_v(\mu, \tau^2$ for heavy tails, $\theta_i \sim \text{Laplace}(\mu, \tau^2)$ gives peak at zero, $\pi\delta_0 + (1 - \pi)\mathcal{N}(\mu, \tau^2)$ for a point mass at zero.
- Decomp $p(\theta, \mu, \tau \mid y) = p(\theta \mid \mu, \tau, y)p(\mu \mid \tau, y)p(\tau \mid y)$.
- For unknown σ^2, default priors are half-Cauchy on the data-level sd, $\sigma \sim \text{Ca}^+(0, C)$ or data-level variance default $p(\sigma^2) \propto 1/\sigma^2$. Use $\tau \sim \text{Ca}^+(0, C)$ or $\tau \sim \text{Unif}(0, C)$ if > 5 reps.

Model Check

- Prior sensitivity analysis.
- Posterior predictive checks using $p(y^{\text{rep}} \mid y) = \int p(y^{\text{rep}} \mid \theta)p(\theta \mid y)d\theta$. Compare replicated data to observed data or compute posterior p-value $p_B = \text{P}\left(T(y^{\text{rep}}, \theta) \geq T(y, \theta) \mid y\right) \approx \frac{1}{J}\sum_{j=1}^J I\left(T(y^{\text{rep},j}, \theta^{(j)}) \geq T(y, \theta^{(j)})\right)$.

Hypothesis Testing

Suppose $y \sim p(y \mid \theta)$.

- All simple ($H_j : \theta = \theta_j\ \forall j$): treat as discrete prior on θ_j: $\text{P}(\theta = \theta_j) = p_j$. Posterior is then $\text{P}(\theta = \theta_j \mid y) = \frac{p_j p(y \mid \theta_j)}{\sum_k p_k p(y \mid \theta_k)}$.
- All composite ($H_j : \theta \in (E_{j-1}, E_j]\ \forall j$): just find area under the curve $\text{P}(H_j \mid y) = \int_{E_{j-1}}^{E_j} p(\theta \mid y)d\theta$.
- Mixed: use $p(H_j \mid y) \propto p(y \mid H_j)p(H_j)$. For simple H_j, $p(y \mid H_j) = p(y \mid \theta_j$ ie $\theta \mid H_j \sim \delta_{\theta_j}$. For composite H_j, $p(y \mid H_j) = \int p(y \mid \theta)p(\theta \mid H_j)d\theta$ is the marginal likelihood and $p(\theta \mid H_j)$ is the prior for θ when H_j is true.

Posterior and Posterior Predictive Propriety

- Tonell's Theorem: if $\mathcal{X}, \mathcal{Y}$ are σ-finite measure spaces and f is nonnegative and measurable, then $\int_{\mathcal{X}}\int_{\mathcal{Y}} f(x, y)dydx = \int_{\mathcal{Y}}\int_{\mathcal{X}} f(x, y)dxdy$.
- Proper prior + discrete data $\implies$ proper posterior.
- Proper prior + continuous data $\implies$ almost surely proper posterior.
- Improper prior $\implies$ prior predictive $p(y) = \int p(y \mid \theta)p(\theta)d\theta$ is improper too.

Bayes Factor

- $p(H_0 \mid y) = \frac{p(y \mid H_0)}{p(y \mid H_0)p(H_0) + p(y \mid H_1)p(H_1)} = \frac{1}{1 + \frac{p(y \mid H_1)p(H_1)}{p(y \mid H_0)p(H_0)}}$.
- $BF(H_1 : H_0) = \frac{p(y \mid H_1)}{p(y \mid H_0)} = \frac{1}{BF(H_0 : H_1)}$.
- Noninformative prior $\implies BF(H_0 : H_1) \to \infty$.
- For $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, $BF(H_0 : H_1) = \frac{p(y \mid \theta_0)}{\int p(y \mid \theta)p(\theta \mid H_1)d\theta}$.

LRT

$H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$. $-2\log\left(\frac{L(\hat{\theta}_{0,MLE})}{L(\hat{\theta}_{MLE})}\right) \overset{d}{\to} \chi^2_{\Delta \dim}$.

Jeffrey-Lindley Paradox

Bayesian and LRT disagree, usually when small effect size, large n, precise H_0, diffuse H_1.