Bayesian Testing

Up until now, we've primarily concerned ourselves with estimation type problems. However, many perform hypothesis tests.

Say, $x \sim N(0, 1)$ there are three types of tests you might consider for testing μ:

1.

2.

3.

In a Bayesian framework, we will use point mass priors for Bayesian hypothesis testing.

Example. Consider testing the hypothesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta_0 \neq \theta_1$. Say you observe data, $\tilde{x} = (x_1, \ldots, x_n)$, where $x_i \sim N(\theta, \sigma^2)$ with σ^2 known.

- **Q:** How would this question be addressed in a classical framework?

- If we want to be Bayesian, we need a prior. Suppose we choose a flat prior, $p(\theta) \propto 1$. Then $p(\theta = \theta_0|\tilde{x}) \sim N(\bar{x}, \sigma^2/n)$. With this distribution, we compute $Pr(H_0|\tilde{x}) =$

- Consider a different prior that places mass on $H_0 : \theta = \theta_0$ which is non-zero. Specifically let $Pr(\theta = \theta_0) =$

- We also need a prior for the alternative space. Let's choose a conjugate prior. Let

- Combining these the prior is:

$$ \tag{1} $$

  where $\delta(\theta = \theta_0)$ is an indicator function for $\theta = \theta_0$ and $p_1(\theta) = N(\mu_1, \tau^2)$.

- Recall $x_i \sim N(\theta, \sigma^2)$. We want to know $Pr(H_0|\tilde{x})$ and $Pr(H_1|\tilde{x})$.

$$ Pr(H_0|\tilde{x}) = \frac{p(\tilde{x}|H_0)p(H_0)}{p(\tilde{x})} $$

  and similarly,

$$ Pr(H_1|\tilde{x}) \propto $$

- So how do we pick $p_1(\theta)$?

  In this course we will use $p_1(\theta) \sim N(\mu_1, \tau^2)$. So how do we pick the parameters of this distribution μ and τ^2?

- Consider the ratio:

$$ \frac{p(\tilde{x}|H_0)}{p(\tilde{x}|H_1)} = \tag{2} $$

  This is known as a Bayes Factor.

- Recall the maximum-likelihood has a related form:

$$ \tag{3} $$

In a likelihood ratio test we compare the difference for specific values of θ that maximize the ratio, whereas the Bayes factor (BF) integrates out the parameter values - in effect averages across the parameter space.

- In this example, let's choose $\mu_1 = \theta_1$ and set $\tau^2 = \psi^2$. Note $\bar{x}$ is a sufficient statistic, so we consider $p(\bar{x}|\theta)$.Then:

$$BF = \frac{\int_{\theta \in H_0} p(\bar{x}|\theta) p_{\theta_0}(\theta) d\theta}{\int_{\theta \in H_1} p(\bar{x}|\theta) p_1(\theta) d\theta} = \frac{\sqrt{n}/\sigma \exp\left(-\frac{(\bar{x}-\theta_0)^2}{2\sigma^2 n}\right)}{1/\sqrt{\sigma^2/n + \psi^2} \exp\left(-\frac{(\bar{x}-\theta_0)^2}{2(\sigma^2/n+\psi^2}\right)}$$

Note that $Pr(H_0|Data) = \left(1 + \frac{1-\pi_0}{\pi_0} BF^{-1}\right)$, (HW problem).

- Example. Let π_0 = 1/2, $\sigma^2 = \psi^2$, $N = 15$, $Z = 1.96$, plugging this all in we get BF = 0.66. This implies:

$$Pr(H_0|\bar{x}) = (1 + .66^{-1})^{-1} =$$

Q: Reject or not?

Q: What is the corresponding p-value here?

Consider the following scenarios with $z = 1.96$.

N	5	10	50	100	1000
$Pr(H_0|\bar{x})$					

In each case the p-value is 0.05. Note that for a given effect size (ψ^2) the Bayes Factor is effect size calibrated. For a given effect size, a p-value goes to zero. Hence the disagreement between 'practical significance' and 'statistical significance'.

- So in this case the relevant question is how to choose ψ^2.

- **Q:** What happens as $\psi^2 \rightarrow \infty$?

  Recall from our example:

$$BF = \left(1 + \frac{N\psi^2}{\sigma^2}\right)^{1/2} \exp\left(-1/2z^2 \left[1 + \frac{\sigma^2}{n\psi^2}\right]^{-1}\right)$$

  so the BF

- Consider two models: $M_1 \& M_2$, each with parameters sets $\Theta^{(M_1)}$ and $\Theta^{(M_2)}$. The Bayes Factor is:

$$BF = \tag{4}$$

  When $p_{M_1}(\Theta^{(M_1)}$ and $p_{M_2}(\theta^{(M_2)}$ are proper, the BF is well defined.

- **Q:** Can you ever specify an improper prior on any of the parameters in $\Theta^{(M_1)}$ and $\Theta^{(M_2)}$?

Bayesian Time Series: State Space models, Filtering, and Sequential Monte Carlo

Another example of latent variable modeling is state space models. A common example is:

(5)

(6)

where y_t is the observed response at time t and x_t is the latent process that evolves in time. This type of model is often thought of as signal plus noise, where the signal is the latent value $\tilde{x}_{1:t}$ and the observed signal *with noise* is $\tilde{y}_{1:t}$. Typically the goal is sequential inference in a Bayesian framework. At time t suppose we know

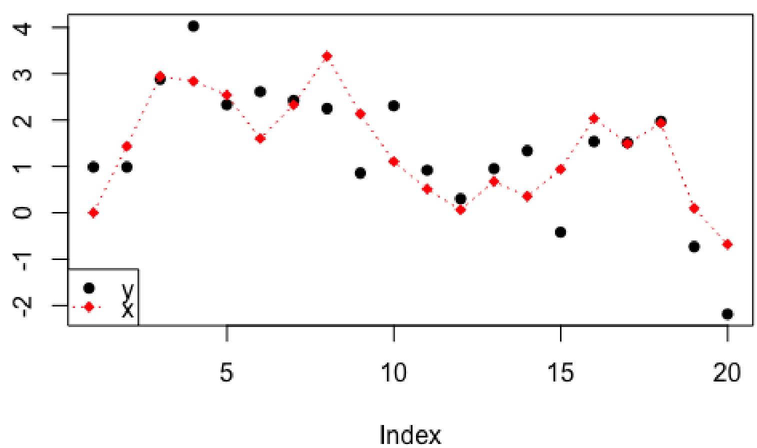

Figure 1: Example of a state space model, where x is the latent space and y is the observed result.

$p(x_t|\tilde{y}_{1:t})$. That is we have a posterior distribution for the unknown latent variable at time t given the observed values from time 1 to t. A typical analysis would involve predicting where the next observed value would be. This takes a couple of steps:

1.

2.

3.

Note the iterative nature of this process, where the posterior at time t becomes the prior for time $t+1$. MCMC methods can be used for fitting these kinds of models, but often require fitting the entire time series in one sweep rather than this iterative process.

A common state space model is :

$$\begin{aligned} \text{observation equation} \qquad y_t &= \mu_t + \nu_t & (7) \\ \text{evolution (state) equation} \qquad \mu_t &= \mu_{t-1} + \omega_t, & (8) \end{aligned}$$

where $\nu_t \sim N(0, V_t)$ and $\omega_t \sim N(0, W_t)$. When W_t and V_t are known this model can be fit using the Kalman filter. That is there exist analytical solutions for the state propagation and predictive distributions, where:

- Posterior for μ_{t-1} : $p(\mu_{t-1}|\tilde{y}_{1:t-1})$

- Prior for μ_t : $p(\mu_t|\tilde{y}_{1:t-1})$

- 1-step forecast: $p(y_t|\tilde{y}_{1:t-1}$

- Posterior for μ_t : $p(\mu_t|\tilde{y}_{1:t})$

Note this process can easily be computed **online.** Online implies the analysis proceeds as the data becomes available in a prospective framework - not retrospectively once all data has been obtained.

Dynamic Linear Models and Sequential Monte Carlo

State space models are frequently used in Bayesian time series (and spatiotemporal) analysis. In particular, dynamic linear models are a popular tool. Consider the following spatiotemporal model:

$$\tilde{y}_t =$$

$$\tilde{\beta}_t =$$

where $\tilde{\epsilon_t} \sim N(0, \sigma^2 H(\phi))$, $H(\phi)$ is a spatial covariance matrix, $\gamma_t \sim N(0, \tau^2)$, and $\tilde{\beta}_t$ are time varying coefficients.

More generally dynamic linear models can be written as:

$$\text{observation equation} \qquad \tilde{y}_t = F\tilde{x}_t + \tilde{\epsilon}_t \tag{9}$$

$$\text{evolution (state) equation} \qquad \tilde{x}_t = G\tilde{x}_{t-1} + \tilde{\gamma}_t. \tag{10}$$

This model incorporates the entire class of ARIMA type time series models in a Bayesian framework.

Dynamic GLMs

Now consider a situation where the response is no longer normal. A common example is count data from the Poisson distribution.

$$\tilde{y}_t = Poisson(\mu_t) \tag{11}$$

$$\log(\mu_t) = F\tilde{x}_t + \tilde{\epsilon}_t \tag{12}$$

$$\tilde{x}_t = G\tilde{x}_{t-1} + \tilde{\gamma}_t. \tag{13}$$

Suppose we still want to do an online analysis. Unfortunately, we don't have analytical solutions to the Bayesian recursive equations.

Q: We can use MCMC, but what variables will we need to track each iteration and what priors are needed?

Another option for fitting this model is sequential Monte Carlo methods. We will see more details about some of these methods next week but a common algorithm is known as the bootstrap filter. Suppose at time t we have a random sample $\{x_t^{(1)}, \ldots, x_t^{(m)}\}$ that approximates $p(x_t|\tilde{y}_{1:t})$, then the algorithm follows as:

1. Sample

2. Weight each particle according to an importance weight

3. Resample each

Similar to MCMC the posterior distribution is approximated by a discrete set of particles.