## Week 10: Part a Oct. 31 -

## 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 $\mu$:

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 $\sigma^2$ 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 $\mu$ and $\tau^2$?

- 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 $\theta$ 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:

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

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

- Example. Let $\pi_0$ = 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 |
|---|---|---|---|---|---|
| $Pr(H_0 \vert \bar{x})$ | | | | | |

In each case the p-value is 0.05. Note that for a given effect size ($\psi^2$) 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 $\psi^2$.

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

  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 $\Theta^{(M_2)}$?