Section 9, 10/28/19

Bayesian Inference (pages 290-293)

In a Bayesian framework, the data comes from distribution $f_{X,\Theta}(x,\theta)$ parameterized by an unknown parameter θ. The unknown parameter θ is treated as a random variable distributed according to some **prior distribution** $f_{\Theta}(\theta)$. We write the joint distribution as the product of the **likelihood** and the prior, that is $f_{X,\Theta}(x,\theta) = f_{X|\Theta}(x|\theta)\, f_{\Theta}(\theta)$. Given observed data, we can think of the likelihood of our parameter, what is called the **posterior distribution** $f_{\Theta|X}(\theta|x) \propto f_{X|\Theta}(x|\theta)\, f_{\Theta}(\theta)$.

Ex. Normal Likelihood

Consider the case of a likelihood distribution which is normal with mean θ. In this case, for convenience, we will write the variance σ^2 instead as the precision $\xi = 1/\sigma^2$. That is our density is

$f(x|\mu,\xi) = \big(\frac{\xi}{2\pi}\big)^{1/2} exp\big(-\frac{1}{2}\xi(x-\theta)^2\big)$

yet we do not know our mean and precision.

Consider the independent priors

$\Theta \sim N(\theta_0, \xi_{prior}^{-1})$
$\Xi \sim \Gamma(\alpha, \lambda)$

Recalling that the likelihood of n observations can be factored, we can then write the posterior distribution as

$$\begin{aligned} f_{\Theta,\Xi|X}(\theta,\xi|x) &\propto f_{X|\Theta,\Xi}(x|\theta,\xi) f_{\Theta}(\theta) f_{\Xi}(\xi) \\ &\propto \xi^{n/2} exp\big(-\frac{\xi}{2}\sum(x_i-\theta)^2\big)\, exp\big(-\frac{\xi_{prior}}{2}(\theta-\theta_0)^2\big)\xi^{\alpha-1} exp(-\lambda\xi) \end{aligned}$$

We can disregard all constants as the rlation is a proportion, not equality. In order to first determine an appropriate estimate for θ, we need to do away with ξ which can be accomplished by "marginalizing" it out through integration. That is,

$f_{\Theta|X}(\theta|x) = \int_0^\infty f_{\Theta,\Xi|X}(\theta,\xi|x)$

Examining the posterior, as a function of ξ is appears to be distributed like a gamma density with parameters $\hat{\alpha} = \alpha + n/2$ and $\hat{\lambda} = \lambda + (1/2)\sum(x_i-\theta)^2$. Thus

$$f_{\Theta|X}(\theta|x) \propto exp\big(-\frac{\xi_{prior}}{2}(\theta-\theta_0)^2\big)\frac{\Gamma(\alpha+n/2)}{[\lambda + 1/2\Sigma(x_i-\theta)^2]^{\alpha_n/2}}$$

This can be solved computationally. Or, if n is large or $\alpha, \lambda, \xi_{prior}$ are small, simplifications show

$f_{\Theta|X}(\theta|x) \propto \big(\Sigma(x_i-\theta)^2\big)^{-n/2}$

This is maximized when $\theta = \bar{x}$.

Hypotesis Testing (page 337)

We have examined confidence intervals before, and here we will show that inverting a confidence set yields a hypothesis test and vice versa. Consider the following motivating example.

Example A

Consider the random samples $X_1, \dots, X_n$ i.i.d. from a $N(\mu, \sigma^2)$ normal distribution. We wish to make the following hypothesis test:

$H_0 : \mu = \mu_0$
$H_A : \mu \neq \mu_0$

We wish to make a claim with certainty level α, that is we wish to be correct $1 - \alpha$ of the time under this model. So, the null hypothesis can be rejected if $|\bar{X} - \mu_0| > x_0$ where x_0 is such that $P(|\bar{X} - \mu_0| > x_0) = \alpha$ when the null hypothesis is true, i.e. $x_0 = \sigma_{\bar{X}} z(\alpha/2)$.

This is equivalent to saying that we have a $100(1 - \alpha)$% confidence interval for μ_0:

$[\bar{X} - \sigma_{\bar{X}} z(\alpha/2), \bar{X} + \sigma_{\bar{X}} z(\alpha/2)]$

The confidence interval consists of all of the values of μ_0 for which the null hypothesis is accepted.

The p-value (pages 334-335)

The p-value is an important concept in the notion of statistical significance testing, in accepting or rejecting a null hypothesis.

The concept of a null hypothesis goes back to the famous statistician Fisher and the "lady tasting tea" experiment. The story goes as follows: The lady in question informed Fisher that she could tell the difference in taste between tea that had had milk added to the cup before or after the tea was poured. Fisher devised an experiment to test this by giving her eight cups of tea, four cups from each category. In this case, it makes sense to assume that a default, the "null distribution", of chance guessing. Fisher was willing to accept the alternative idea that she could tell the difference only if she correctly identified all 8 cups. That is, if she could, this result was unlikely enough to reject the idea the idea that she was just guessing. Here, the null distribution of cups is known and the alternative distribution is unknown and arbitrary.

The idea is that we assume a null distribution to be true. If the likelihood of the data under this assumed truth is small, then this gives us reason to doubt and reject our assumption of the truth. The notion of "small" requires the selection of a significance level α prior to analysis of the data. This choice is arbitrary, usually 0.05 or 0.10 by convention.

Once we have a defined, or approximated as often is the case, the null distribution, given the the data we can calculate the probability of a result being *more extreme* than what was observed if the null distribution is true. This is the definition of a **p-value**. If the p-value is lower than the significance level α, then our observation is extreme enough to warrant rejecting the null. Note, the p-value is **not** the probability that the null hypothesis is true, a common misconception.

Bayesian Testing

Suppose we have observed data and hypothesize two potential distributions from which the data could have come. Often the distribution is the same and we hypothesize two potential sets of parameters H_0 and H_1. In this case, our posterior probabilities are

$P(H_0|x) \quad \text{and} \quad P(H_1|x)$

By Bayes law

$$P(H_0|x = \frac{P(H_0,x)}{P(x)}) = \frac{P(x|H_0P(H_0))}{P(x)}$$

Thus, we can express the ratio of the posterior probabilities as

$$\frac{P(H_0|x)}{P(H_1|x)} = \frac{P(H_0)}{P(H_1)}\frac{P(x|H_0)}{P(x|H_1)}$$

the product of the priors and likelihoods. Once we define our priors, this ratio is a defined number. In order to choose a hypothesis, it makes sense to choose H_0 if the ratio is greater than 1, i.e. the likelihood of H_0 is greater than that of H_1.

$$\frac{P(H_0|x)}{P(H_1|x)} > 1$$

Since the ratio of the priors is some constant, this is equivalent to

$$\frac{P(x|H_0)}{P(x|H_1)} > c$$

for some threshold value c based on our prior beliefs. That is $P(x|H_0) > cP(x|H_1)$. It turns out that the choice of c balances the tradeoff between Type 1 and Type 2 errors.

Type 1 Error: P(reject $H_0|H_0$)

Type 2 Error: P(accept $H_0|H_1$)