# Stat 330 - Tutorial 12



This week's tutorial is review for the exam. I will focus on material from after the second midterm. Let's start with Bayesian inference, since we haven't talked about it yet.

## Bayesian Statistics

The biggest ~~change~~ new idea in Bayesian Statistics is that we now think of the parameter as being random. This has some implications for what we should do with data.

Let $\pi$ be the distribution of $\theta$. We call this the prior distribution for $\theta$.
The distribution of $\theta$ given the data is called the posterior distribution of $\theta$. We usually also write $\pi$ for the posterior distribution.
The posterior distribution of the parameter given our sample turns out to be the central object for Bayesian inference. Let's look at how to calculate the posterior.

$$\pi(\theta \mid \underline{X}) = \frac{f(\theta, \underline{X})}{f(\underline{X})} = \frac{f(\underline{X} \mid \theta) \cdot \pi(\theta)}{\int f(\underline{X} \mid \theta)\pi(\theta)\,d\theta} = \frac{\mathcal{L}(\theta; \underline{X}) \cdot \pi(\theta)}{\int \underline{\qquad} \, d\theta}$$

~~The denominator is the~~ The numerator is the likelihood for $\theta$ based on our sample, times the prior for $\theta$. The denominator is the integral of the numerator w.r.t. $\theta$. In particular, the denominator does not depend on $\theta$ (i.e. it is constant), and we can calculate the denominator from the numerator. Because of this, it is common to call the denominator a proportionality constant and write

$$\pi(\theta \mid \underline{X}) \propto \mathcal{L}(\theta; \underline{X}) \cdot \pi(\theta)$$

If we ever need to recover this proportionality constant, we can just integrate.

e.g. 1 Let $\theta \sim Beta(\alpha, \beta)$ and $X \mid \theta \sim Bin(n, \theta)$. Find the posterior distribution of $\theta \mid X$.

Note first that the beta distribution has density $\pi(\theta) = \dfrac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$, where $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

~~This gives us the prior~~. The beta distribution is also only defined for $\theta \in (0, 1)$.

②

The likelihood is

$$\mathcal{L}(\theta; X) = \binom{n}{X} \theta^X (1-\theta)^{n-X}$$

Multiplying these ~~by~~ together, we get

$$\mathcal{L}(\theta; X)\,\pi(\theta) = \binom{n}{X} \theta^X (1-\theta)^{n-X} \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

$$= \frac{1}{B(\alpha,\beta)} \binom{n}{X} \theta^{X+\alpha-1} (1-\theta)^{n-X+\beta-1}$$

$$\propto \theta^{X+\alpha-1}(1-\theta)^{n-X+\beta-1} \quad \text{everything we dropped is constant in } \theta$$

We're almost done, but this expression ~~is~~ is only proportional to the posterior, not equal to the posterior. ~~To~~ To get equality, we need to ~~m~~ get the proportionality constant. Fortunately, our expression looks a lot like the density of a beta distribution, and we know that the beta density must equal 1.

$$\int_0^1 \frac{\theta^{X+\alpha-1}(1-\theta)^{n-X+\beta-1}}{B(X+\alpha, n-X+\beta)}\, d\theta = 1$$

$$\int_0^1 \theta^{X+\alpha-1}(1-\theta)^{n-X+\beta-1}\, d\theta = B(X+\alpha, n-X+\beta)$$

Therefore, if we divide our expression by $B(X+\alpha, n-X+\beta)$, we get a valid density. This is just the density of the beta distribution, so $\theta | X \sim \text{Beta}(X+\alpha, n-X+\beta)$. Alternatively, we could have ~~s~~ saved a few steps by recognizing that $\mathcal{L}(\theta; X)\cdot\pi(\theta)$ is proportional to the kernel of a beta distribution (the part of the density that depends on $\theta$).

Side Note: From the data, we have $X$ ~~h~~ successes & $n-X$ failures. When combining this information with the prior, we keep the same type of distribution, but increase the first parameter by $X$ & the second parameter by $n-X$. We can think of the parameters in the ~~bet~~ beta distribution as representing the number of heads and tails seen respectively. This means that the prior for $\theta$ can be thought of as representing ~~$\alpha$ heads and $\beta$~~ $\alpha$ successes & $\beta$ failures before we see the data.

//

We now have some idea how to calculate the posterior distribution of $\theta$. ~~How~~ This is only useful if it helps us do inference. For this course, we focus on point estimation & interval estimation. You can also do Bayesian hypothesis testing, just not here. See Stat 460 for more Bayesian goodness.

## Bayesian Point Estimation

The posterior distribution measures our understanding of the parameter $\theta$. How should we use this to guess the true value of $\theta$? It turns out that there are multiple ways. The two we focus on are the posterior mean and the posterior mode. These are both what they sound like: the posterior mean is $E(\theta|\underline{X})$, and the posterior mode is the maximizer of the posterior distribution.

### e.g. 1 cont.

We know that the posterior distribution of $\theta$ is $Beta(X+\alpha, n-X+\beta)$, with density

$$\pi(\theta|X) = \frac{\theta^{X+\alpha-1}(1-\theta)^{n-X+\beta-1}}{B(X+\alpha, n-X+\beta)}$$

The posterior mean is

$$E(\theta|X) = \int_0^1 \frac{\theta \cdot \theta^{X+\alpha-1}(1-\theta)^{n-X+\beta-1}}{B(X+\alpha, n-X+\beta)}\, d\theta$$

$$= \frac{1}{B(X+\alpha, n-X+\beta)} \int_0^1 \theta^{X+\alpha+1-1}(1-\theta)^{n-X+\beta-1}\, d\theta$$

$$= \frac{B(X+\alpha+1, n-X+\beta)}{B(X+\alpha, n-X+\beta)} \int \frac{\theta^{X+\alpha+1-1}(1-\theta)^{n-X+\beta-1}}{B(X+\alpha+1, n-X+\beta)}\, d\theta$$

$$= \frac{\Gamma(X+\alpha+1)\Gamma(n-X+\beta)}{\Gamma(n+\alpha+\beta+1)} \Bigg/ \frac{\Gamma(X+\alpha)\Gamma(n-X+\beta)}{\Gamma(n+\alpha+\beta)}$$

$$= \frac{(X+\alpha)}{(n+\alpha+\beta)}$$

The posterior mean is $\hat{\theta} = \frac{(X+\alpha)}{(n+\alpha+\beta)}$

4

To get the posterior mode, let's drop the proportionality constant and take the log.

Let $\tilde{\pi}(\theta|X) = \theta^{X+\alpha-1}(1-\theta)^{n-X+\beta-1} \propto \pi(\theta|X)$

$\log[\tilde{\pi}(\theta|X)] = (X+\alpha-1)\log(\theta) + (n-X+\beta-1)\log(1-\theta)$

$$\nabla_\theta \log[\tilde{\pi}(\theta|X)] = \frac{X+\alpha-1}{\theta} - \frac{n-X+\beta-1}{1-\theta} = 0$$

$$\Rightarrow (X+\alpha-1)(1-\tilde{\theta}) - \tilde{\theta}(n-X+\beta-1) = 0$$

$$X+\alpha-1 - X\tilde{\theta} - \alpha\tilde{\theta} + \tilde{\theta} - n\tilde{\theta} + X\tilde{\theta} - \beta\tilde{\theta} + \tilde{\theta} = 0$$

$$2\tilde{\theta} - \alpha\tilde{\theta} - \beta\tilde{\theta} - n\tilde{\theta} + X + \alpha - 1 = 0$$

$$\tilde{\theta}(\alpha+\beta+n-2) = X+\alpha-1$$

$$\tilde{\theta} = \frac{X+\alpha-1}{\alpha+\beta+n-2}$$

The posterior mode is $\tilde{\theta} = \dfrac{X+\alpha-1}{\alpha+\beta+n-2}$

//

## Bayesian Interval Estimation

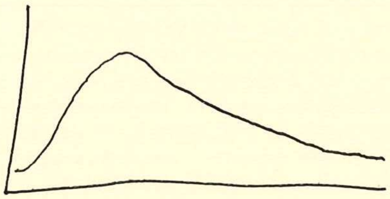
Bayesian interval estimates are great. You get to interpret them the way you wanted to interpret confidence intervals in Stat 270. Our parameter has a distribution, so we can just construct an interval that contains 95% of the probability under this distribution. When we say "the probability that $\theta$ is in our interval is 95%", we mean this as a probability statement about $\theta$.

Like with point estimation, it turns out that there are multiple ways to do Bayesian interval estimation. Two major ways are the percentile method and the highest posterior density (HPD) method. Intervals obtained by either of these methods are called credible intervals.

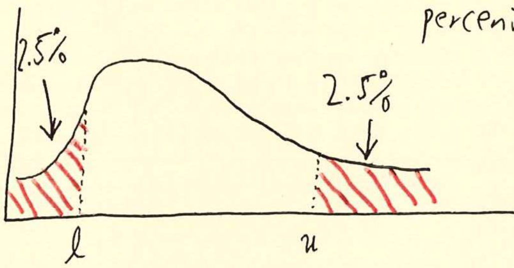
The calculations required to obtain percentile or HPD credible intervals are hard, and people usually get them from R. However, the ideas are pretty straightforward. We will compare them using pictures.

e.g. 2

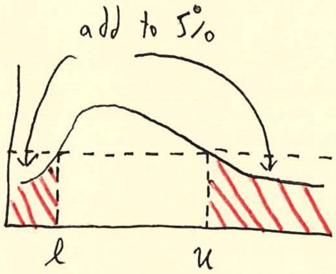
Consider a Bayesian statistical model with the following posterior density:



Sketch the 95% percentile and HPD credible intervals.




percentile ~~m~~ method




HPD method

Note that in the HPD method, we don't know ~~the~~ how much prob. is in the lower or upper tail individually, we just know that these probabilities sum to 5%.