Lecture 10

Informative priors for the Beta-Binomial model

Consider the following data set. There are 6,115 mothers. Each mother had greater than or equal to 13 children. We only consider their first 12 children (thus each mother has 13 children in this data set). We now count the number of boys for each mother.

#Boys	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
X	3	24	104	286	670	1033	1343	1112	829	478	181	45	7	6115

How do we model this data $\mathcal{F}$? This example is beyond the scope of the course. For instance, $X \sim Bin(12, 50\%)$. Its turns out the sex ration is not even. The probability of having a boy is closer to 51%, not 50%. That difference is real. So let's examine the model $X \sim Bin(12, 51\%)$.

#Boys	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
X	3	24	104	286	670	1033	1343	1112	829	478	181	45	7	6115
Bin(12, 0.51)	1	12	72	259	628	1085	1367	1266	854	410	152	26	2	6115

It seems like the binomial model just cannot capture the tail ends. Let us do the Beta-Binomial model now. How do we fit a Beta-Binomial? We know that $n = 12$, but what is alpha and beta? We fit alpha and beta with the maximum likelihood and find $\alpha_{MLE} = 34$ and $\beta_{MLE} = 32$. We get,

$$X \sim BetaBinom(12, 34, 32)$$

$$E[X] = \frac{12 \times 34}{34 + 32} = 0.515$$

51.5% is the published average. We now get this table given this function,

#Boys	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
X	3	24	104	286	670	1033	1343	1112	829	478	181	45	7	6115
Bin(12, 0.51)	1	12	72	259	628	1085	1367	1266	854	410	152	26	2	6115
BetaBinom(12, 34, 32)	2	23	105	311	656	1036	1258	1182	854	462	178	44	5	6115

The beta binomial model fits better to human birth data, it is saying that every mother herself is a draw from the beta β. That is, the mothers θ value is drawn from the distribution $P(\theta)$

$$P(\theta) = Beta(34, 32)$$

$$Q[\theta, 0.005] = 36\%$$

$$Q[\theta, 0.995] = 67\%$$

Back to the curriculum... What about the following problem. You see data for n Bernoulli trials. What if you want to know the the next future n_* trials you have not seen? This problem is called the "prediction" problem, i.e. forecasting. In science there are generally two goals. Explaining phenomena, which means finding a model $\mathcal{F}$ and estimating its parameters, and predicting the future values of the phenomena. They are related. We have mainly been studying the first goal, we're going to venture into the second goal now. Can we use what we know about θ to talk about what the future will bring. Consider the following if θ is known:

$$P(X_*|X = x) = Binom(n_*, \hat{\hat{\theta}}_{MLE})$$

The problem is θ is never known. Is using the MLE a reasonable idea? Yes, and people do do this, but we can do better. The problem with the above is $\hat{\theta}_{MLE}$ is not θ and there is uncertainty in its estimation that is not being accounted for. We know that with n large, the MLE is approximately normally distributed. We can use this, but if n is small, it won't be accurate. So... Bayesian statistics to the rescue!

$$P(X_*|X = x) = \int_\Theta P(X_*, \theta|X)d\theta = \int_\Theta P(X_*, \theta|X)P(\theta|X)d\theta$$

The posterior predictive distribution is defined as the third expression, $\int_\Theta P(X_*, \theta|X)P(\theta|X)d\theta$. It is defined as a mixture/compound random variable. To reiterate, $P(X_*|\theta)$ is the likelihood, and $P(\theta|X)$ is the posterior.

$$\int_\Theta P(X_*, \theta|X)P(\theta|X)d\theta = BetaBinom(n_*, \alpha + x, \beta + n - x)$$

$$P(X_*, \theta|X) = Bin(n_*, \theta)$$

$$P(\theta|X) = Beta(\alpha + x, \beta + n - x)$$

Example: We see $n = 10$ at bats for a new baseball player and he gets $x = 6$ hits. Assuming each at bat is $iidBern(\theta)$, what is the probability he will have $x_* = 17$ hits in the next $n_* = 32$ at bats? Assume a uniform prior.

$$P(\theta) = Beta(1, 1)$$

$$P(X_*|X=6) = BetaBinom(32, 1+6, 1+4)$$

$$P(X_*=17|X=6) = \frac{\binom{32}{17}}{B(7,5)} B(24,20) = dbetabinomial(17,32,7,5)$$

$$dbetabinomial(X_*,\ n_*,\ \alpha+x,\ \beta+n-x)$$

$$P(X_* \leq 17|X=6) = \Sigma_{y=0}^{17} \frac{\binom{32}{y}}{B(7,5)} B(y+7, 32-y+5) = pbetabinomial(17,32,7,5)$$

We started with Laplace's prior but we found chinks in the armor, that is, we have to assume a success and a failure, and the n is greater than or equal to 2. We looked at Haldane's prior but we found that the prior distribution is improper because there exist no successes or failures and n = 0. Lets try to find a more appropriate prior. Back to probability land, let X, Y be continuous random variables where f_X and is known, and $Y = t(X)$ where t is a known invertible function. We want to derive f_Y using f_x and t.