341 Lec10

Andrew Claros

Birth ratios

- 6115 mothers, each has ≥ 13 children, but only consider the first 12 children. Count the number of boys for each mother.

# of Boys	0	1	2	3	4	5	6	7	8	9	10	11	12	total
x	3	24	104	206	670	1033	1343	1112	829	478	181	45	7	6115
Binom	1	12	72	259	628	1085	1367	1266	854	410	152	26	2	6115
BetaBin	2	23	105	311	656	1036	1258	1182	854	462	178	44	5	6115

Binomial

- Modeling this data ($\mathcal{F}$) is beyond the course. If the mother has a baby, what is the probability it will have a boy? $X \sim \text{Binomial}\left(12,\, 50\%\right)$ does not accurately capture the data. $\mathbb{P}\left(\text{boy}\right) \approx 51$.

Beta Binomial

- Solvining alpha and beta through maximum likelihood:

$$n = 12, \alpha_{MLE} = 34, \beta_{MLE} = 32 \to X \sim \text{BetaBinomial}\left(12,\, 34,\, 32\right), \mathbb{E}\left[X\right] = 12\frac{34}{34+12} = 0.511 = 51\%$$

- The BetaBinomial fits better.

- $\mathbb{P}\left(\theta\right) = \text{Beta}\left(34,\, 32\right), \text{Quantile}\left[\theta,\, 0.5\%\right] = 36\%, \text{Quantile}\left[\theta,\, 99.5\%\right] = 67\%$

Generalized:

- After seeing n bernoulli trials, what is the *next future* n_* trials not seen?

- Conside: $\mathbb{P}\left(x_*|X = x\right) \stackrel{?}{=} \text{Binomial}\left(n_*,\, \hat{\hat{\theta}}_{MLE}\right)$ It is reasonable to use MLE but the problem is that $\hat{\hat{\theta}}_{MLE}$ is not θ

- With $\gg 0, \theta$ is approximately normally distributed.

- For small n we need Bayesian statistics

Posterior predictive distribution

- $\mathbb{P}\left(X_*|x\right) = \int_\Theta \mathbb{P}\left(X_*, \theta|x\right) d\theta = \int_\Theta \mathbb{P}\left(X_*|\theta, x\right)\mathbb{P}\left(\theta|x\right) d\theta \to \int_\Theta \mathbb{P}\left(X_*\right)\mathbb{P}\left(\theta|x\right) d\theta$ (a mixture/compound distribution)

- For $\mathcal{F} : \text{Bin}(n, \theta) : \int \underbrace{\mathbb{P}\left(X_*|\theta\right)}_{\text{Bin}(n,\theta)} \underbrace{\overbrace{\mathbb{P}\left(\theta|x\right)}^{\text{prior } \mathbb{P}(\theta)\text{beta}}}_{\text{Beta}(\alpha+x,\, \beta+n-x)} d\theta = \text{BetaBinomial}\left(n_**,\, \alpha + x,\, \beta + n - x\right)$

- $\mathbb{P}\left(\theta\right) \stackrel{x}{\to} \mathbb{P}\left(\theta|x\right)$ and $\mathbb{P}\left(x_*\right) \stackrel{x}{\to} \mathbb{P}\left(X_*\right)|x$

Example:

- $n = 10$ at bats for a new baseball player and $x = 6$ hits assuming each at bat is $\overset{iid}{\sim}$ Bernoulli$\,(\theta)$ what is the probability he will hace $x_* = 17$ hits in the next $n_* = 32$ at bats? Assume uniform prior.

$\mathbb{P}\left(X_*|x\right) = \text{BetaBinomial}\left(32,\ 7,\ 5\right), \mathbb{P}\left(X_* = 17|x = 6\right) = \frac{\binom{32}{17}}{\beta(7,5)}\beta(24, 20) = \text{dbetabinom}(17, 32, 7, 5)$

Probability 17 or less hits on the next 32 at bats?

- $\mathbb{P}\left(X_* < 17|x = 6\right) = \sum_{y=0}^{17} \frac{\binom{32}{y}}{\beta(7,5)}\beta y + 7, 32 - y + 5 = \text{pbetabinom}(17, 32, 7, 5)$

Probability theory

- Let X, Y be continuous rv's where f_x, f_X ar known and $y = t(X)$ where t is a known unvertible function. Derive f_Y using f_X and t
- $\mathbb{P}\left(x \in A\right) = \mathbb{P}\left(y \in B\right)$
- If A,B small $\mathbb{P}\left(x \in A\right) \approx f_x(x)|dx|, \mathbb{P}\left(y \in B\right) \approx f_y|dy| \Rightarrow f_x(x)|dx| \approx f_y|dy|_y(y) = f_x(x)|\frac{dx}{dy}|$
  - $f_y(y) \overset{y=t(x), x=t^{-1}(y)}{=} f_x(t^{-1}(y))|\frac{d}{dy}[t^{-1}(y)]$