From lec 9:

$$\boxed{1-\theta = \frac{\beta}{\alpha+\beta}}$$

$$n \lim_{\alpha\to\infty} \frac{\theta(1-\theta)}{(\theta+1-\theta)^2} = n\,\theta(1-\theta) \text{ which is the same variance as } \text{Bin}(n,\theta)$$

$$\text{VAR}[X] = n\frac{\alpha\beta}{(\alpha+\beta)^2}\;\frac{\alpha+\beta+n}{\alpha+\beta+1} = \underbrace{n\,\theta(1-\theta)}_{\text{VAR Binomial}}\;\frac{\alpha+\beta+n}{\alpha+\beta+1}$$

overdispersion $\leftarrow \in (1, n)$

More flexible model, Beta Bin allows for extreme outcomes.

Consider the following data set - 6115 mothers, each had 13 or more children and we only consider their first 12 children; thus each has 12 in this data set. (Not considering children past 12, 13, 14, 15 etc)

Look at 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
Binomial predictions	1	12	72	259	628	1085	1367	1266	854	410	152	26	2	6115
Beta Bin	2	23	105	311	656	1036	1258	1182	854	462	178	44	5	6115

← Fits Better!

How do we model this theta (?)? This is beyond scope of course.

Example $X \sim \text{Bin}(12, 50\%)$, turns out sex ratio is not even

P(boy) is closer to 51.2% not 50%. The difference is real.

Lets examine $X \sim \text{Bin}(12, 51.2\%)$

How do we fit a Beta Binomial? We know $n = 12$, what is α and β? We fit α, β with MLE and find $\alpha_{MLE} = 34$ $\beta_{MLE} = 32$.

So now we have $X \sim \text{BetaBin}(12, 34, 32)$. $E[X] = \frac{n\alpha}{\alpha+\beta} = \frac{12 \cdot 34}{34+32}$

$= .515$

BetaBin, has similar mean but higher variance, allows for better model of the tails.

$P(\theta) = \text{Beta}(34, 32)$ $\quad Q[\theta, .5\%] = 36\%$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad Q[\theta, 99.5\%] = 67\%$

<u>Back to Baye's</u> What about the following problem. You see data for n Bernoulli trials. What if you want to know about the next future n^* trials you have not seen?

$\frac{0}{1}\ \frac{1}{2}\ \cdots\cdots\ \frac{1}{n}$ | $\frac{\ }{1}\ \frac{\ }{2}\ \cdots\cdots\ \frac{\ }{n^*}$

$X = \#$ of 1's above | $X_* = ?$ ↑ future, have not seen. how many successes will I see.

The problem is called the "prediction" problem ("forecasting")

In science two goals: (1) Explaining phenomena, which means finding a model and estimating its parameters (θ) and (2) predicting future values of the phenomena. They are related.

Can we use what we know about theta to answer questions about the future?

Consider example. $P(X_* \mid X = x)$ ~~this is called~~ $\overset{?}{=} \text{Bin}(n_*, \hat{\theta}_{MLE}) = \text{Bin}(n_*, \frac{x}{n})$

If theta known $P(X_* \mid X = x) = \text{Bin}(n_*, \theta)$ but theta is never known!

Is using the MLE a reasonable? Yes. We can do better. Problem with the above. $\hat{\theta}_{MLE}$ is not θ and there is uncertainty in its estimation that is not being accounted for.

We know with n large, the MLE is approximately normally distributed. But if n is small it won't be accurate. Bayesian statistics to the rescue.

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

(will be a sum if Θ is discrete)

If θ is known X doesn't give you any more info so drop it.

posterior predictive distribution

$$= \int_\Theta P(X_* | \theta, X)\, P(\theta | X)\, d\theta = \int_\Theta \underbrace{P(X_* | \theta)}_{\text{likelihood}}\, \underbrace{P(\theta | X)}_{\text{posterior}}\, d\theta$$

It is a mixture/compound.

For

F: Bin(n, θ) prior $P(\theta)$ = beta

$$\int_\Theta \underbrace{P(X_* | \theta)}_{\text{Bin}(n_*, \theta)}\, \underbrace{P(\theta | X)}_{\text{Beta}(\alpha + x,\ \beta + n - x)}\, d\theta = \text{BetaBin}(n_*, \alpha + x, \beta + n - x)$$

Previously

$$P(\theta) \xrightarrow{x} P(\theta | X) \text{ but also} \ldots \overset{\text{inference}}{P(X_*)} \xrightarrow{x} \overset{\text{prediction}}{P(X_* | X = x)}$$

As your ideas about theta get better, your ideas about future data get better.

$$\int_\Theta p(X|\theta)\, p(\theta)\, d\theta \longrightarrow \int_\Theta P(X_* | \theta)\, P(\theta | X)\, d\theta$$

Example: A new baseball player has $n=10$ at bats and he gets $X=6$ hits assuming each at bat is iid $\text{Bern}(\theta)$, what is the probability he will have $X_* = 17$ hits in the next $n_* = 32$ at bats? Assume uniform prior.

$$P(\underset{\overset{\|}{6}}{X_*} \mid X) = \text{BetaBin}(32, 7, 5)$$

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

what is the probability he gets 17 or less hits on the next 32 at bats?

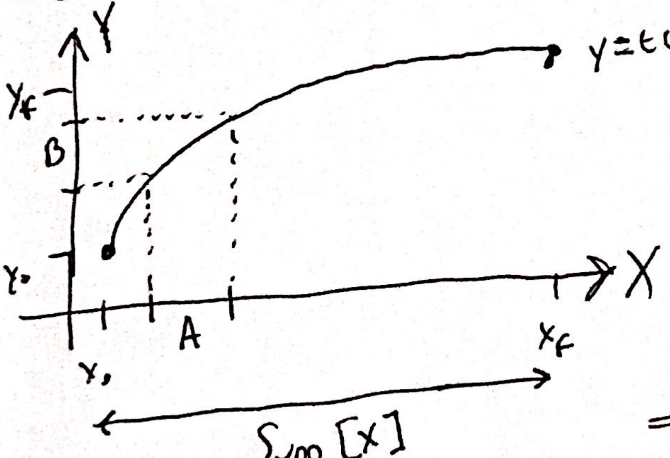
$$P(X_* \leq 17 \mid X=6) = \sum_{y=0}^{17} \frac{\binom{32}{y}}{B(7,5)} B(y+7, 32-y+5) = \text{pbetabinomial}(17, 32, 7, 5)$$

We will come back to this later. Go to probability land again:

Let X, Y be cts r.v.'s such that $Y = t(X)$ where t is a known invertible function.

where f_X is known.

We want to derive f_Y using f_X and t.



$P(X \in A) = P(Y \in B)$, If A, B small

$P(X \in A) \approx f_X(x)\,|dx|$

$P(Y \in B) \approx f_Y(y)\,|dy|$

Therefore these two are equivalent

$$f_X(x)\,|dx| = f_Y(y)\,|dy| \qquad y = t(x),\; x = t^{-1}(y)$$

$$\Rightarrow f_Y(y) = \frac{f_X(x)\,|dx|}{|dy|} =$$

change of variables formula for densities.

$$\Rightarrow f_Y(y) = f_X(t^{-1}(y)) \left| \frac{d}{dy}\left[t^{-1}(y)\right] \right| .$$