Consider the following dataset. There are 6,115 mothers. Each mother had >12 children and we only consider their first 12 children (thus each mother has 12 children in this dataset). We now count the # of boys for each mother:

# Boys	0	1	2	3	4	5	6	7	8	9	10	11	12	tot
x	3	24	104	286	670	1033	1343	1112	829	478	181	45	7	6115
Binomial pred's	1	12	72	259	628	1085	1367	1266	854	410	152	26	2	6115
Betabinom preds	2	23	105	311	656	1036	1258	1182	854	462	178	44	5	6115

How do we model this data (curly F). This example is beyond the scope of the course. E.g. X ~ Bin(12, 50%). It turns out, the sex ratio is not even: P(boy) is closer to 51.1% (not 50%). That differenc is real. So let's examine the model X ~ Bin(12, 51.1%)

How do we fit a betabinomial? We know n=12. What is alpha and beta? We fit the alpha and beta with maximum likelihood and find alphaMLE = 34 and betaMLE = 32. So now we have X ~ BetaBinom(12, 34, 32).
E[X] = 12 * 34 / (34 + 32) = 0.515 ~ 51.1% (the published avg).

The betabinomial model fits better to human birth data.

$P(\theta) = \text{Beta}(34, 32)$ $\quad Q[\theta, 0.5\%] = 36\%$

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

Back to the curriculum... 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 haven't seen?

$\underset{1}{0} \; \underset{2}{1} \; \ldots \; \underset{n}{1}$ | $\underset{1}{\_} \; \underset{2}{\_} \; \ldots \; \underset{n_*}{\_}$

x = # 1's above $\qquad X_* = ?$

past $\qquad$ future

This problem is called the "prediction" problem ("forecasting"). In science there are generally two goals: (1) explaining phenomena which means finding a model curlyF and estimating θ, its parameters and (2) predicting the future values of the phenomena. They are related. Consider:

$$P(X_* \mid X = x) \overset{?}{=} \text{Bin}(n_*, \hat{\theta}_{MLE}) = \text{Bin}\left(n_*, \frac{x}{n}\right)$$

$\overset{\text{if } \theta \text{ known}}{\longrightarrow} \; \| \; \text{Bin}(n_*, \theta)$ but θ is never known!

Is using the MLE a reasonable idea? Sure... but we can do better! The problem with the above is thetahatMLE is not theta and there is uncertainty in its estimation that is not being accounted for. We know 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!

$$P(X_* \mid X) = \int_{\Theta} P(X_*, \theta \mid x)\, d\theta = \int_{\Theta} P(X_* \mid \theta, x)\, P(\theta \mid x)\, d\theta$$

If θ is known, X doesn't give you any more information.

posterior predictive distibution, a mixture / compound rv

$$= \int_{\Theta} \overbrace{P(X_* \mid \theta)}^{\text{likelihood}} \underbrace{P(\theta \mid x)}_{\text{posterior}}\, d\theta$$

for $\mathcal{F}$: Bin(n, θ), prior P(θ) beta

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

$$P(\theta) \xrightarrow{X} P(\theta \mid x) \quad \text{but also...} \quad P(X_*) \xrightarrow{X} P(X_* \mid X)$$

$$P(X_*) = \int_{\Theta} P(X_* \mid \theta) P(\theta)\, d\theta \longrightarrow \int_{\Theta} P(X_* \mid \theta) P(\theta \mid x)\, d\theta$$

Let's see a concrete example. We see n = 10 at bats for a new baseball player and he gets x = 6 hits. Assuming each at bat is iid Bern(θ), what is the probability he will have x_* = 17 hits in the next n_* = 32 at bats? Assume a uniform prior. $P(\theta) = \text{Beta}(1, 1)$

$$P(X_* \mid X = 6) = \text{BetaBinomial}(32, 1 + 6, 1 + 4)$$

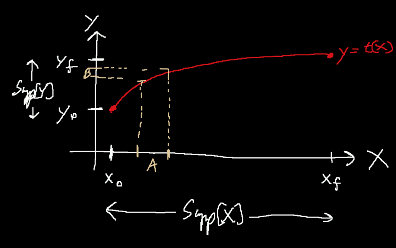
$$P(X_* = 17 \mid X = 6) = \frac{\binom{32}{17}}{B(7, 5)} B(24, 20) = \text{dbetabinom}(17, 32, 7, 5)$$

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

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

$$= \text{pbetabinom}(17, 32, 7, 5)$$

Back to probability land... Let X, Y be continuous rv's where f_X is known and Y = t(X) where t is a known invertible function. We want to derive f_Y using f_X and t.



$P(X \in A) = P(Y \in B)$

If A, B small,

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

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

$\Downarrow$

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

$\Downarrow$

$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$

This is called the change of variables formula for densities: $\boxed{f_Y(y) = f_X(t^{-1}(y)) \left| \frac{d}{dy}\left[t^{-1}(y)\right] \right|}$