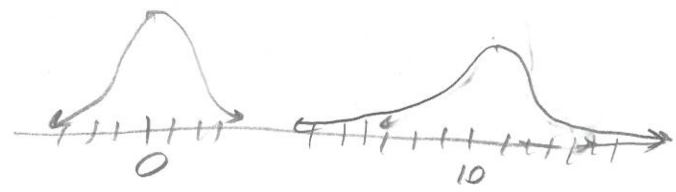
Math 341 Lec 9 3/7/19

Let's first talk about mixture distributions...

$$X \sim \begin{cases} N(0, 1^2) & \text{wp } \frac{1}{2} \\ N(10, 2^2) & \text{wp } \frac{1}{2} \end{cases}$$



What is PDF of X? Use Law of Total Prob.

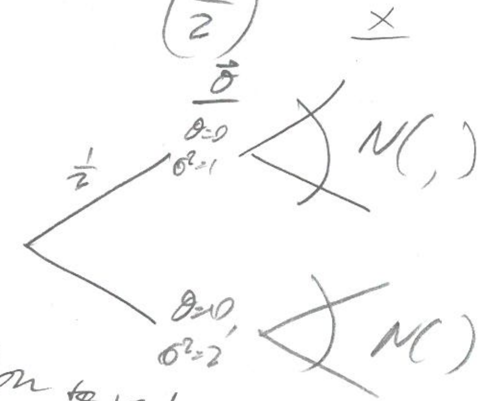
$$P(X) = \sum_{\theta \in \Theta} P(X|\theta) P(\theta) = P(X|\theta=1, \sigma^2=1^2)\, P(\theta=1, \sigma^2=1^2) + P(X|\theta=10, \sigma^2=2^2)\, P(\theta=1, \sigma^2=2^2)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \left(\frac{1}{2}\right) + \frac{1}{\sqrt{8\pi}} e^{-\frac{1}{4}(x-10)^2} \left(\frac{1}{2}\right)$$

component

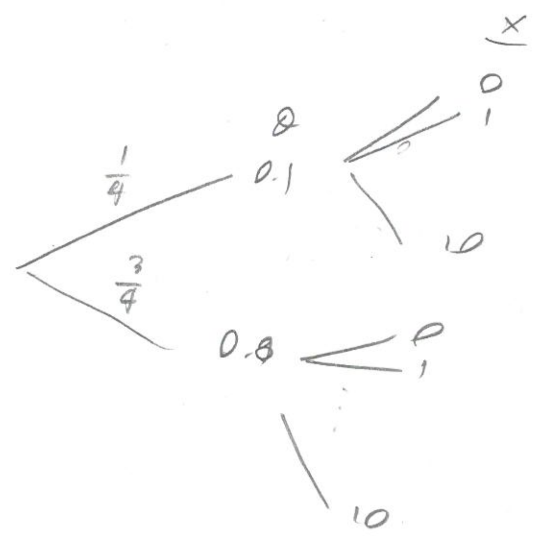
$$P(x) = \int_{\Theta} P(x|\theta) P(\theta)\, d\theta$$

Mixture distr. model components mixing proportions



How to realize an x

$$X \sim \begin{cases} Bin(10, 0.1) & \text{wp } \frac{1}{4} \\ Bin(10, 0.8) & \text{wp } \frac{3}{4} \end{cases}$$



$$P(x) = \frac{1}{4}\binom{10}{x}(0.1)^x(0.9)^{10-x} + \frac{3}{4}\binom{10}{x}(0.8)^x(0.2)^{10-x}$$

Have we seen this before? $P(\theta|x) = \dfrac{P(x|\theta)\,P(\theta)}{P(x)}$

$$P(x) = \int_{\Theta} P(x|\theta)\,P(\theta)\,d\theta$$

model components are the different likelihoods

mixing proportions is the prior

~~If $P(\theta) = \text{Beta}(\alpha,\beta)$, $P(x|\theta) = \text{Binom}(n,\theta)$~~
~~To prove $P(\theta|x) = \text{Beta}(x+\alpha, n-x+\beta)$ we used the "fixed" trick~~
~~$P(\theta|x)$~~

Let's return to the denominator of the posterior and just compute the denominator

$P(\theta) = \text{Beta}(\alpha,\beta)$, $P(x|\theta) = \text{Binom}(n,\theta)$

$$P(x) = \int_0^1 \binom{n}{x}\theta^x(1-\theta)^{n-x}\,\frac{1}{B(\alpha,\beta)}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}\,d\theta = \frac{\binom{n}{x}}{B(\alpha,\beta)}\,B(x+\alpha, n-x+\beta) = \text{BetaBinom}(n,\alpha,\beta)$$

AKA the "overdispersed binomial" we will see why soon...

$X \sim \text{BetaBinom}(n,\alpha,\beta)$

$\text{Supp}(X) = \{0, 1, \ldots, n\}$

Param Space $\underbrace{n \in \mathbb{N}}_{\text{valid param space of likelihood model (binomial)}},\ \overbrace{\alpha > 0,\ \beta > 0}^{\text{valid param space of mixing distr (beta)}}$

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

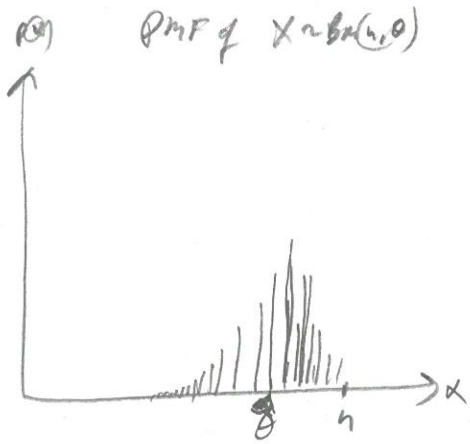
$$\text{Var}(X) = \ldots = n\frac{\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

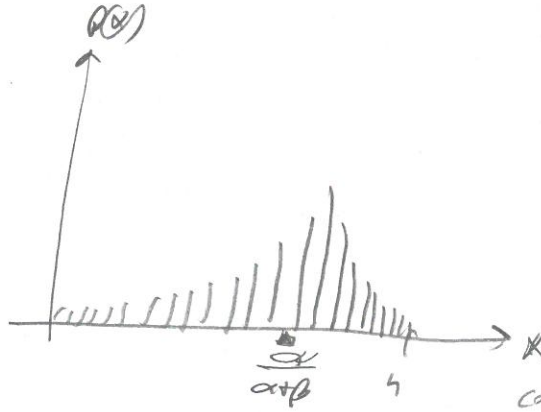
Let $\theta := \frac{\alpha}{\alpha+\beta} \Rightarrow \theta\alpha + \theta\beta = \alpha \Rightarrow (\theta - 1)\alpha = -\theta\beta \Rightarrow \beta = \alpha\frac{1-\theta}{\theta}$

$E(X) = n\theta$ just like binomial! What happens to variance as we let $\alpha \to \infty$ while pegging $\frac{\alpha}{\alpha+\beta} = \theta$ i.e. keep the ratio fixed

$$\lim_{\alpha\to\infty} Var(X) = \lim n \frac{\alpha\left(\alpha\frac{1-\theta}{\theta}\right)\left(\alpha + \alpha\frac{1-\theta}{\theta} + n\right)}{\left(\alpha + \alpha\frac{1-\theta}{\theta}\right)^2\left(\alpha + \alpha\frac{1-\theta}{\theta} + 1\right)} = n \lim \frac{\alpha^2\left(\frac{1-\theta}{\theta}\right)}{\alpha^2\underbrace{\left(1+\frac{1-\theta}{\theta}\right)^2}_{\frac{1}{\theta}}} \frac{\alpha\left(1+\frac{1-\theta}{\theta}\right)+n}{\alpha\left(1+\frac{1-\theta}{\theta}\right)+1}$$

$$= n\theta(1-\theta) \lim_{\alpha\to\infty} \underbrace{\frac{\frac{\alpha}{\theta}+n}{\frac{\alpha}{\theta}+1}}_{\to 1} = n\theta(1-\theta) = \text{Variance of the binomial!}$$





$Var(X) = n\theta(1-\theta)$

$Var(X) = n\theta(1-\theta)\frac{\alpha+\beta+n}{\alpha+\beta+1}$ can control variance (within bounds)

$\alpha, \beta \to \infty$ at same rate

$\Rightarrow$ Binomial

$\Rightarrow Beta(\alpha, \beta) \to Deg(\theta)$

Let's make an example of the BetaBinomial

Gender Birth Data. 6,115 females with ≥ 13 children, consider their first 13 children

# 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 predicted	1	12	72	259	628	1085	1367	1266	854	410	132	26	2	6115
BetaBinomial predicted	2	23	105	311	656	1036	1258	1182	854	462	178	44	8	6115

$\theta = 0.511$ birth prob. of boy $\neq 0.5$ (Bach 241)

model:

$X \sim \text{Bin}(n = 12, \theta = 0.511)$

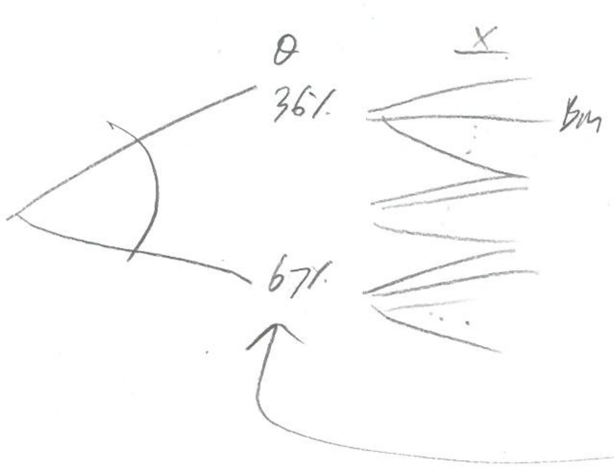
Tails are ~~too~~ small!! How can we make tails larger? Overdispense!

let $X \sim \text{BetaBin}(n = 12, \alpha, \beta)$ and select most probable α, β

using $\hat{\alpha}_{MLE}, \hat{\beta}_{MLE}$

$\hat{\alpha}_{MLE} = 34$, $\hat{\beta}_{MLE} = 32$

In the mixture dist.... $P(\theta) = \text{Beta}(34, 32)$, $E(\theta) = 0.515 \approx 0.511$



$SE[\theta] = 0.061$

$Q[\theta, 0.005] = 0.36$

$Q[\theta, 0.995] = 0.67$

99% of mass

Back to our [illegible] problem ... $\mathcal{F}$ = Binomial

If $P(\theta) = Beta(\alpha,\beta)$

$$\underbrace{X_1, \ldots, X_n}_{n}, X^* \overset{iid}{\sim} Bern(\theta) \qquad P(X^*|x) = Bern\left(\frac{x+\alpha}{n+\alpha+\beta}\right)$$

predict on one future r.v

What if..

$$X_1, \ldots, X_n, \underbrace{X_1^*, \ldots, X_{n^*}^*} \overset{iid}{\sim} Bern(\theta)$$

predict on n^* future r.v.'s Call the whole thing X^*

$$P(X^*|\theta) = Binom(n^*, \theta)$$

But we don't know θ! We again want the posterior predictive distr:

$$P(X^*|x) = \int_{\Theta} \underset{Binomial(n^*,\theta)}{P(X^*|\theta)} \; \underset{Beta(x+\alpha,\, n-x+\beta)}{P(\theta|x)} \, d\theta$$

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

$$= BetaBinom(n^*, x+\alpha, n-x+\beta)$$

$P(\theta) \xrightarrow{x} P(\theta|x)$ inference is updated

$P(X^*) \xrightarrow{x} P(X^*|x)$ predictions are updated