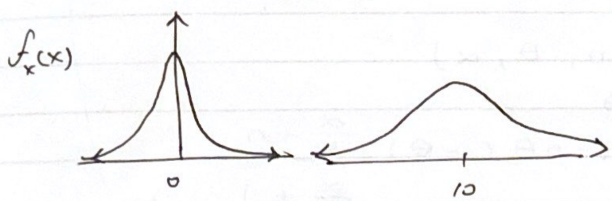
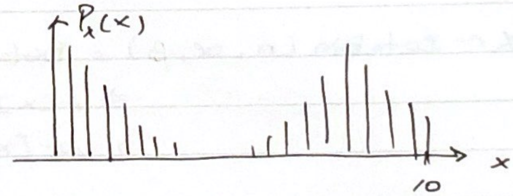
Bayesian

03/03/2020.
class #9.

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

$$x \sim \begin{cases} \text{Bin}(10, 0.1) & \text{wp } \frac{1}{4} \\ \text{Bin}(10, 0.6) & \text{wp } \frac{3}{4} \end{cases}$$





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

$$: \sum_{\theta \in \Theta} P(x \mid \theta)\, p(\theta)$$

$$P(\theta|x) = \frac{\overset{\text{assumed with } \mathcal{F}}{\uparrow}\; P(x|\theta)\,P(\theta)}{P(x)}$$

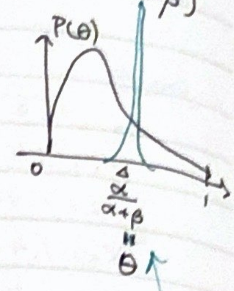
$$P(x) = \int_{\Theta} \underbrace{P(x|\theta)}_{\text{likelihood model}} \underbrace{P(\theta)}_{\text{mixing proportions (prior).}} d\theta$$

$\mathcal{F}$: Binomial

$P(\theta) = \text{Beta}(\alpha, \beta)$

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

$P(x) = ?$



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

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

as a result of mixing

$$= \binom{n}{x}\frac{B(x+\alpha, n-x+\beta)}{B(\alpha,\beta)} = X \sim \text{Beta Binomial}(n, \alpha, \beta)$$

$$E[X] = \sum_{x=0}^{n} x\, p(x) = \dots = n\frac{\alpha}{\alpha+\beta}$$

$$\text{Supp}[X] = \{0, 1, \dots, n\}$$

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

param Space : $n \in \mathbb{N}$

$\alpha > 0$

$\beta > 0$

A different parametization

Let

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

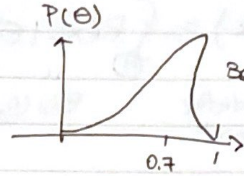
$$X \sim \text{BetaBin}(n, \alpha, \beta) = \text{BetaBin}(n, \theta, \alpha)$$

$$\Rightarrow E[X] = n\theta$$

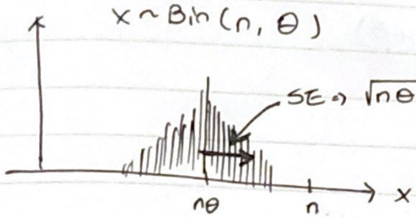
$$\Rightarrow \text{Var}[X] = \dots = n\theta(1-\theta)\frac{\frac{\alpha}{\theta}+n}{\frac{\alpha}{\theta}+1}$$

$$\lim_{\alpha\to\infty} \text{Var}[X] = n\theta(1-\theta)\lim_{\alpha\to\infty}\cancel{\frac{\frac{\alpha}{\theta}+n}{\frac{\alpha}{\theta}+1}}^{\,1} = n\theta(1-\theta)$$

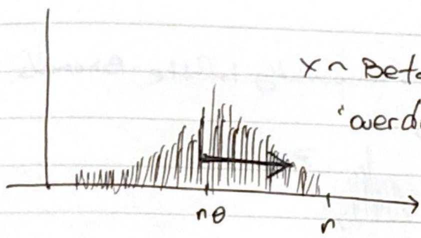
$\theta = 0.7 = \frac{\alpha}{\alpha+\beta}$, $\alpha = \cancel{0.7}^{\,3.5} \Rightarrow \beta = \cancel{0.3}^{\,1.5}$



$x \sim Bin(n, \theta)$



$x \sim Beta(n, \alpha, \beta)$
"overdispersed binomial"



$\rightarrow Var = n\theta(1-\theta)\frac{\frac{\alpha}{\theta}+n}{\frac{\alpha}{\theta}+1} = \frac{\frac{3.5}{0.7}+n}{\frac{3.5}{0.7}+1} = \frac{5+n}{6}.$

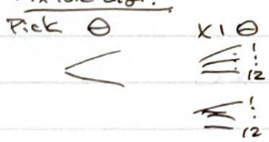
Birth Gender Data Example

$P(male) = 0.51$
6115 women will ≥ 12 children.

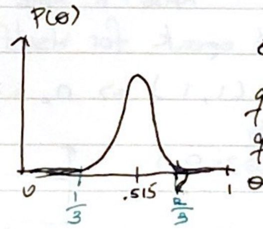
Model 1 : $Bin(12, 0.51)$
Model 2 : $BetaBin(12, 34, 32)$
$\rightarrow \alpha_{MLE} = 34$ → greater than β because .51
$\rightarrow \beta_{MLE} = 32$

# male	0	1	2	3	4	5	6	7	8	9	10	11	12	Σ
F_1 x	3	24	104	286	670	1033	1343	1112	829	478	181	45	7	6115
Model 1 prediction	1	12	72	259	628	1085	1367	1266	854	410	152	26	2	6115
Model 2 prediction $= \hat{F}_2$	2	23	105	311	656	1036	1288	1152	854	462	178	44	5	6115

Mixture dist.



this is saying every women have diff. θ



$E[\theta] = 0.515$
$qbeta(.005, 34, 32) = .36$
$qbeta(.995, 34, 32) = .67$

$\hat{F}$, Binomial fixed n

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

Imagine n_* future observations where $n_* \geq 1$
If θ was known
$x_* \sim Bin(n_*, \theta)$

x_* be # of sucess of n_* future observations.

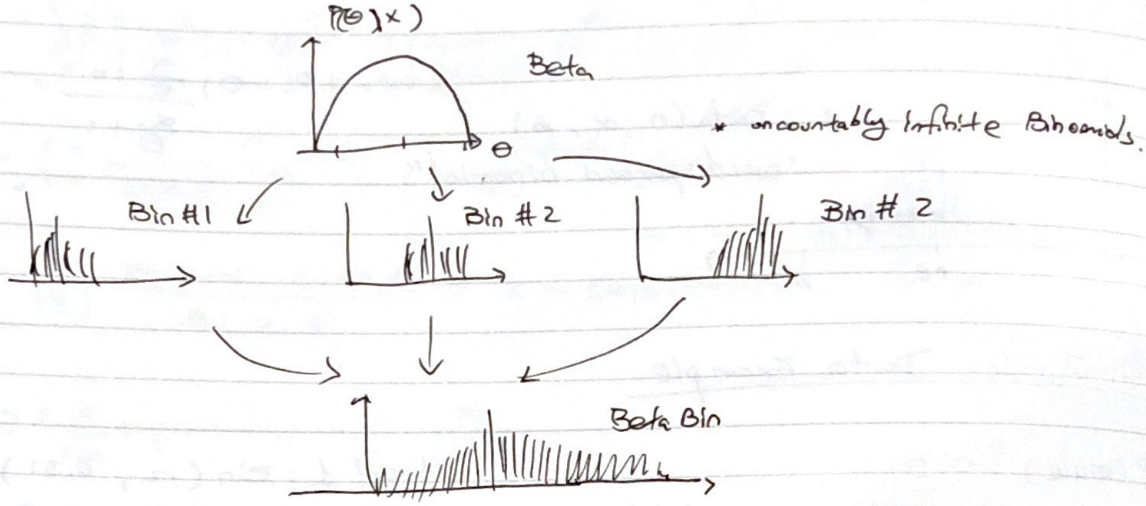
In real life θ is unknown let's use Bayesian Inference. We obtain $P(\theta|X)$

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

'posterior predictive distribution'

$\qquad\qquad\qquad$ " $\text{Bin}(n_*, \theta)$ $\qquad$ " $\text{Beta}(\alpha+x,\ n-x+\beta)$

$$= \text{BetaBin}(n_*,\ \alpha+x,\ \beta+n-x)$$



$= U(0,1)$

$P(\theta) = \text{Beta}(1,1) = \text{Laplace}$

$P(\theta) = \text{Beta}(\tfrac{1}{2},\tfrac{1}{2}) = \text{Jeffrey's}$

Haldane Prior (1932).

$P(\theta) = \text{Beta}(0,0)$ not a legal dist.

Haldane retorts "I don't care!"

$\Rightarrow P(\theta|x) = \text{Beta}(x, n-x)$ will be "proper" if $x \neq 0$ and $x \neq n$

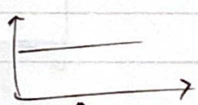
$\hookrightarrow \hat{\theta}_{MMSE} = \frac{x}{n} = \hat{\theta}_{MLE}$ (no shrinkage)

Objectivist: the data must speak for itself.

$P(\theta) = U(0,1) = \text{Beta}(1,1) \Rightarrow n_0 = 2 \Rightarrow \rho = \frac{\alpha+\beta}{\alpha+\beta+n} > 0$

$\Downarrow$

$E[\theta] = 0.5$

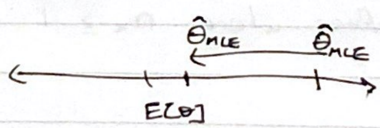


Informative Priors

$\theta \sim \text{Beta}(\alpha, \beta)$

α, β are "large" relative to n.

$E[\theta] = \frac{\alpha}{\alpha+\beta}$



$\rho = \frac{\alpha+\beta}{\alpha+\beta+n}$ is large.