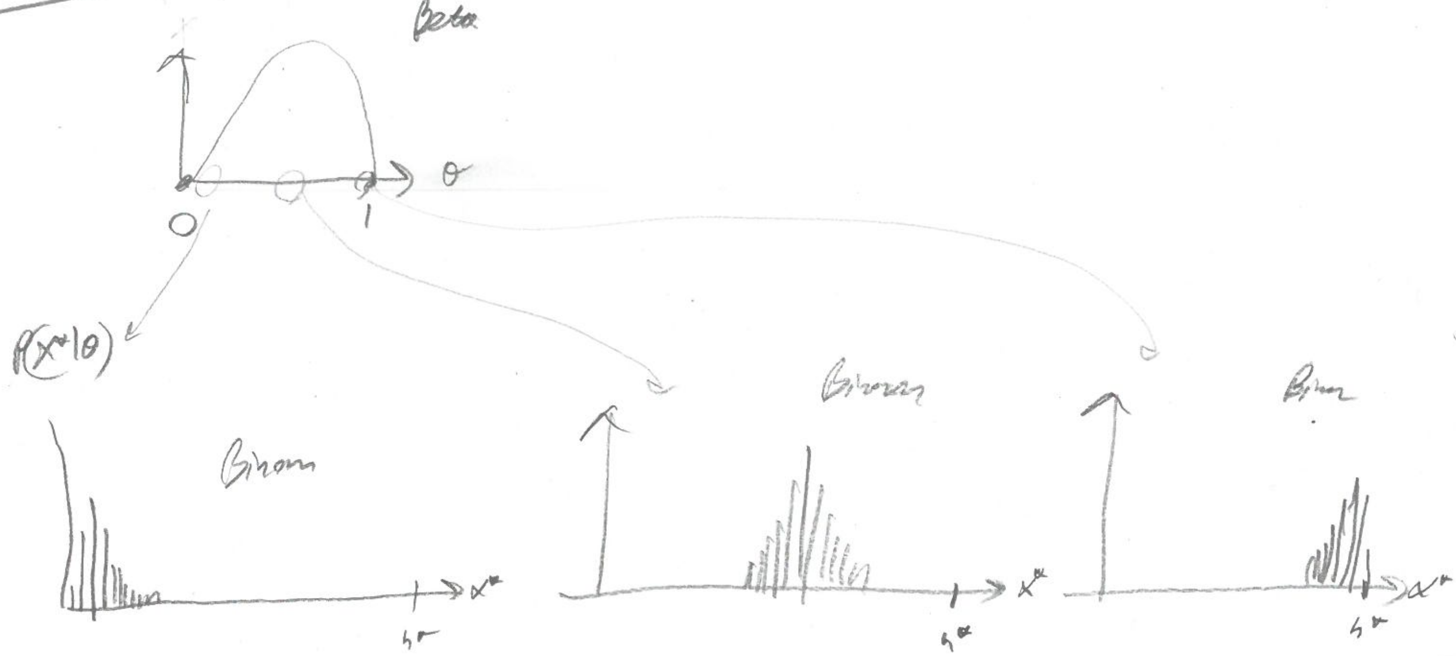
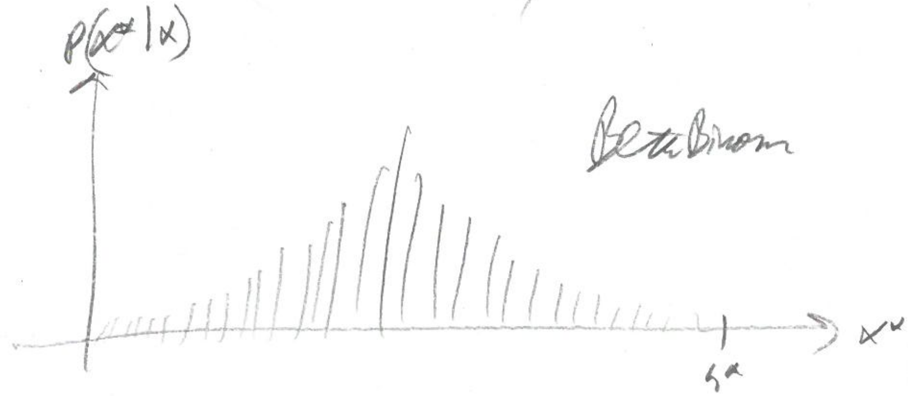
Math 341 Lec 10 3/12/19



If you average over all θ's



See 19 heads out of n = 26. What is the prob in the next 10 flips of 6 H? $P(\theta)=U(0,1)$

$P(X^*|X) = \text{BetaBinom}(10, 20, 8)$

$P(X^*=6|X) = \text{dbetabinom}(6, 10, 20, 8) = .17$

Prob of 6 or less H?

$P(X^* \le 6|X) = \text{pbetabinom}(6, 10, 20, 8) = .33$

Improper Priors

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

$n_0 = \alpha + \beta$ # pseudosamples. If $P(\theta) = U(0,1) \Rightarrow n_0 = 2$. What if we want prior to have no weight at all?

$P(\theta) = \text{Beta}(0,0)$ but this is not a true Beta distr! $\alpha \not> 0, \beta \not> 0$

$\Rightarrow$ "Improper Prior", "Haldane prior" (1932)

Complete ignorance $n_0 = 0$. Unknown if success / failure is even possible.

However, $p(\theta | x) = \text{Beta}(x, n-x)$ which is proper if $x > 0$ and $x < n$.

$$\hat{\theta}_{MMSE} = \frac{x}{n} = \hat{\theta}_{MLE}$$

Information?

This is okay to use as long as posterior is proper.

Informative priors.

Imagine you are trying to infer a baseball player's true batting ability, θ, i.e. prob(hit). Assume $\mathcal{F}$ = Binomial

$\Rightarrow$ all at bats independent, prob(hit) shared but unknown.

$$\hat{\theta}_{MLE} = \frac{x}{n}$$

x ← # Hits

n ← # At Bats

$\hat{\theta}_{MLE}$ performs poorly at inferring θ if n is small.

$n = 3$, $x = 2$ $\quad \hat{\theta}_{MLE} = .667$ Is this possible??? No!

$= \hat{\theta}_{MMSE}$ under ignorance

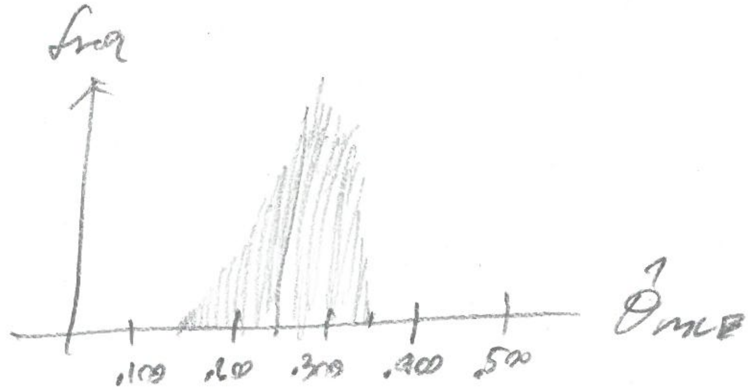
Highest: Ty Cobb .366

If we use Bayesian estimator with $p(\theta) = U(0,1) \Rightarrow \hat{\theta}_{MMSE} = 0.6$ ~~not better~~

Solution? Shrink towards previous data!! All time avg batting avg is $\approx .260$.

Let's design a prior with high informativeness and high shrinkage $\Rightarrow \alpha, \beta$ large.

Method: look at previous data. Subset where each player has $n \geq 500$ at bats. Plot all $\hat{\theta}_{MLE}$'s:



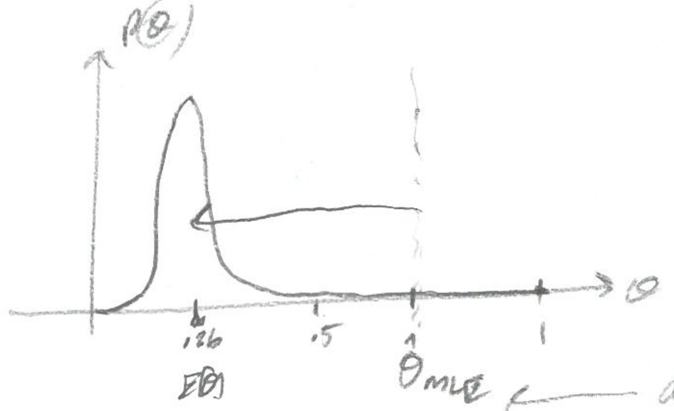
Now... fit a ~~Beta~~ distr. to it. How? Calc. $\hat{\alpha}_{MLE}$, $\hat{\beta}_{MLE}$.

$\Rightarrow P(\theta) = \text{Beta}(78.7, 227.8)$ $\Rightarrow E[\theta] = 0.26$ $\Rightarrow n_0 = 303.5$ empirical Bayes!

For $n=3$ at bats $\quad \rho = \frac{303.5}{303.5+3} = 99\%$

$\hat{\theta}_{MMSE} = (1\%)(.667) + (99\%)(.26) = 0.263$

Why is this smart?



← do you really believe it??

No... it is an artifact of randomness in small datasets. Shrink it back.

Is it possible this guy ~~bats~~ .667? Yes... but the ~~tradeoff~~ is we need a lot of data to prove it.

Even with a strong prior. If n is large $\Rightarrow$ $\rho \to 0$ and $\hat{\theta}_{MLE}$ is $\approx$ ~~our~~ estimate.

---

3 topics in prob.

$\text{Odds}(A) := \text{Odds}(A, A^c)$ e.g. $P(A) = 0.2$

$$:= \frac{P(A)}{P(A^c)} = \frac{P(A)}{1-P(A)} = \frac{0.2}{0.8} = \frac{1}{4} = 1:4$$ usually this

$\text{OddsAgainst}(A) = \text{Odds}(A)^{-1} = 4:1$

$P(A) \in [0,1]$ but...

$\text{Odds}(A) \in [0, \infty)$

$\text{Odds}(A, B) = \frac{P(A)}{P(B)}$ prob. ratio of two events, Imagine you want to compare two θ's:

$$P(\theta = \theta_a \mid X) = \frac{P(X \mid \theta = \theta_a)\, P(\theta = \theta_a)}{P(X)}$$

$$P(\theta = \theta_b \mid X) = \frac{P(X \mid \theta = \theta_b)\, P(\theta = \theta_b)}{P(X)}$$

$$\Rightarrow \frac{P(\theta = \theta_a \mid X)}{P(\theta = \theta_b \mid X)} = \frac{P(X \mid \theta = \theta_a)}{P(X \mid \theta = \theta_b)} \frac{P(\theta = \theta_a)}{P(\theta = \theta_b)}$$

posterior odds prior odds

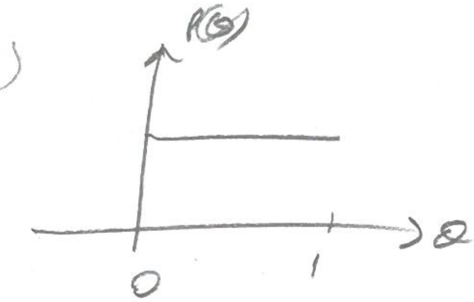
$$\Rightarrow \text{Odds}(\theta_a, \theta_b \mid X) = \frac{P(X \mid \theta = \theta_a)}{P(X \mid \theta = \theta_b)} \text{Odds}(\theta_a, \theta_b)$$

$$\Rightarrow \frac{\text{Odds}(\theta_a, \theta_b \mid X)}{\text{Odds}(\theta_a, \theta_b)} = \frac{P(X \mid \theta = \theta_a)}{P(X \mid \theta = \theta_b)}$$

$$\text{Odds}(\theta_a, \theta_b) \xrightarrow{X} \text{Odds}(\theta_a, \theta_b \mid X)$$

$1:1$ $\qquad\qquad$ $5:1$

F = Binomial $\quad P(\theta) = U(0,1) = \text{Beta}(1,1)$



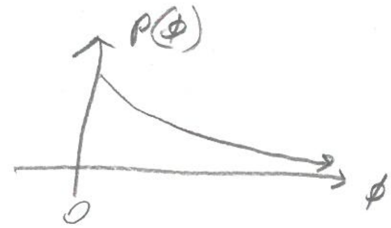
prin of indifference

But is it indifferent on the odds scale? Let $\phi(\theta) = \dfrac{\theta}{1-\theta} \in [0,\infty)$

$P(\theta) = U(0,1) \Rightarrow P(\phi(\theta)) = U(0,\infty)$? No since $U(0,\infty)$ doesn't make sense!!

$\Rightarrow$ the principle of indifference has a big problem!

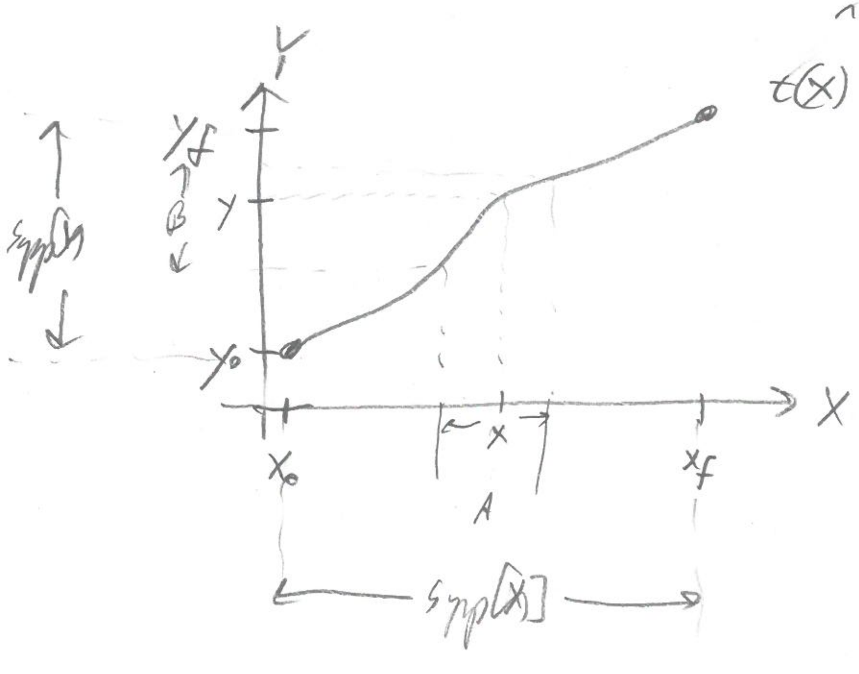
To see the problem clearly we need $P(\phi)$.



Math 621 (not covered here).

Imagine cont. r.v.'s X, Y with densities f_X, f_Y. f_X known

Let $Y = t(X)$ where t is an known invertible function. Find f_Y.



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

If A, B small...

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

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

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

$$y = t(x) \Rightarrow x = t^{-1}(y)$$

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