Lecture 7

- Binomial Distribution $X \sim \text{Binomial}(n, \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ ← counts successes so Σx_i is not needed

$\theta \sim U(0,1)$ * let $\mathcal{F}$ be = Binom $\underbrace{\theta \sim U(0,1)}_{\text{hidden}}$ or $\theta \sim Beta(1,1)$

$$P(\theta|X) = \frac{P(X|\theta, n)\,P(\theta)}{P(X)} = \frac{\theta^x(1-\theta)^{n-x}\cdot 1}{\int_0^1 \theta^x(1-\theta)^{n-x}\,d\theta} = \frac{\theta^{(x+1)-1}(1-\theta)^{(n-x+1)-1}}{\int_0^1 \theta^{(x+1)-1}(1-\theta)^{(n-x+1)-1}\,d\theta}$$

$P(\theta) = 1$

$P(X|\theta, n) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ ← constant; n is fixed

$P(X) = \int_\Theta P(X|\theta)P(\theta)\,d\theta = \int_0^1 \binom{n}{x}\theta^x(1-\theta)^{n-x}\,d\theta$ ← constant

- $P(\theta|X) = \left[Beta\left(\underset{\alpha'}{x+1}, \underset{\beta'}{n-x+1}\right)\right]$ posterior parameter

*More generally if θ is not uniform

$\theta \sim Beta(\alpha,\beta)$

- $\mathcal{F}$ = Binomial ; Prior := $\theta \sim Beta(\alpha,\beta)$ so now:

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

Note: $\binom{n}{x}$ constant; $\frac{1}{B(\alpha,\beta)}$ constant; exponents can add

$$P(\theta|X) = \frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}{\int_0^1 \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}\,d\theta} = \left[Beta\left(\underset{\alpha'}{x+\alpha}, \underset{\beta'}{n-x+\beta}\right)\right]$$

- Note: Normally : $\theta \xrightarrow{\text{data } X} \theta|X$

Now : $Beta(\alpha,\beta) \xrightarrow[\text{data } X]{} Beta(\alpha',\beta')$ the beta is "conjugate" for binomial likelihood

Point estimate

- Mode: $\hat{\theta}_{MAP} = \hat{\theta}_{MLE} = \text{Mode}[\theta|X] = \frac{\alpha'-1}{\alpha'+\beta'-2} = \frac{(x+\alpha)-1}{(n+\alpha+\beta)-2}$ (Max a posteriori)

- Mean: $\hat{\theta}_{MASE} = \arg\min\{E[(\theta-\hat{\theta}_{MASE})^2]\} = E[\theta|X] = \frac{\alpha'}{\alpha'+\beta'} = \frac{(x+\alpha)}{(n+\alpha+\beta)}$ (min. squared error loss)

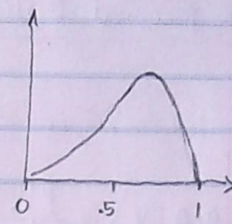
- Median: $\hat{\theta}_{MAE} = \arg\min\{E[|\theta-\hat{\theta}_{MAE}|]\} = \text{Med}[\theta|X] = $ qbeta(0.5, α', β') ; qbeta(0.5, $x+\alpha$, $n-x+\beta$) } code to be run in 'R' program (min. absolute error loss)

$\theta \sim Beta(1,1)$

ex: $\mathcal{X} \sim \text{Binomial}(10, \theta)$ $\quad \theta \sim U(0,1)$; $\quad X = 7$

so: $\theta | X \sim Beta(7+1, 10-7+1)$

$\theta | X \sim Beta(8, 4)$



pt est: $\hat{\theta}_{MAP} = \hat{\theta}_{MLE} = \frac{7}{10} = \frac{8-1}{8+4-2} = \frac{7}{10} = 0.7$

$\hat{\theta}_{MASE} = E[\theta|X] = \frac{8}{8+4} = \frac{8}{12} = 0.66\ldots$

$\hat{\theta}_{MAE} = med[\theta|X] = qbeta(0.5, 8, 4) = 0.676\ldots$

Prediction: The new distribution given $\{ X^* \mid n^* = 1 \text{ (one more trial)} \}$ $\quad supp[X^*] = \{0,1\}$

so: $P(X^*|X) = \int_\Theta \underbrace{P(X^*|\theta)}_{(\text{Bernoulli } \theta)} \underbrace{P(\theta|X)}_{(Beta(x+\alpha, n-x+\beta))} d\theta = \int_0^1 \left[\theta^{x^*}(1-\theta)^{1-x^*}\right]\left[\frac{1}{B(x+\alpha, n-x+\beta)} \cdot \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}\right] d\theta$

$P(X|z) = P(X,\theta|z)$

$= P(X|\theta z)P(\theta|z)$

$= P(X|\theta)P(\theta|z)$

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

$$\left[P(X^*|X) = \frac{B(x^*+x+\alpha,\ 1-x^*+n-x+\beta)}{B(x+\alpha,\ n-x+\beta)} \right]$$

Must be written as a bernoulli

$n^* = 1$

$$P(X^*=1 | X) = \frac{B(x+\alpha+1,\ n-x+\beta)}{B(x+\alpha,\ n-x+\beta)} = \frac{\frac{\Gamma(x+\alpha+1)\,\cancel{\Gamma(n-x+\beta)}}{\Gamma(n+\alpha+\beta+1)}}{\frac{\Gamma(x+\alpha)\,\cancel{\Gamma(n-x+\beta)}}{\Gamma(n+\alpha+\beta)}} = \frac{\frac{x+\alpha}{n+\alpha+\beta}\,\frac{\cancel{\Gamma(x+\alpha)}}{\cancel{\Gamma(n+\alpha+\beta)}}}{\frac{\cancel{\Gamma(x+\alpha)}}{\cancel{\Gamma(x+\alpha+\beta)}}}$$

gamma function

$$\left[P(X^*=1 | X) = \frac{x+\alpha}{n+\alpha+\beta} \right]$$

$X^* = 1 \mid X \sim Bern\left(\frac{x+\alpha}{n+\alpha+\beta}\right)$

$\sim Bern\left(\hat{\theta}_{MASE} = E[\theta|X]\right)$

Note, we won't use $\hat{\theta}_{MLE}$ b/c $\hat{\theta}_{MLE} = \frac{x}{n}$ and if $\underset{\hat{\theta}_{MLE} = 0}{x=0}$; $\underset{\hat{\theta}_{MLE} = 1}{x=n}$: too certain

- Statements from Posterior / Prior distributions (θ)

$$\underset{\text{"pseudocounts"}}{\overset{\text{Prior}}{Beta(\alpha, \beta)}} \xrightarrow{X} Beta(\underbrace{X+\alpha}_{\text{\# of data successes}}, \underbrace{n-X+\beta}_{\text{\# of data failures}})$$

$U(0,1)^{**}$

OR $Beta(\underset{\substack{\uparrow \\ \text{prior success} \\ \text{"pseudosuccess"}}}{\alpha=1}, \underset{\substack{\uparrow \\ \text{prior failure} \\ \text{"pseudofailure"}}}{\beta=1}) \xrightarrow{X} Beta(X+1, n-X+1)$

** $U(0,1)$ is a statement of belief not devoid of info.

** Principle of indifference: $E(\theta) = \frac{\alpha}{\alpha+\beta} = \frac{1}{1+1} = \frac{1}{2} = 0.5$

that is, we believe one failure + success occurred.

$E(\theta)$ is the assumed prior where $E(\theta|X)$ is our mean value after data occurs

* Note: $\hat{\theta}_{MASE} = E[\theta|X] = \frac{\alpha+X}{\alpha+\beta+n} = \frac{\alpha}{\alpha+\beta+n} + \frac{X}{\alpha+\beta+n} = \frac{\alpha\,(\alpha+\beta)}{(\alpha+\beta)(\alpha+\beta+n)} + \frac{n \quad X}{(\alpha+\beta+n)\; n}$

$\left(\frac{\alpha+\beta}{\alpha+\beta}\right) \quad \left(\frac{n}{n}\right)$

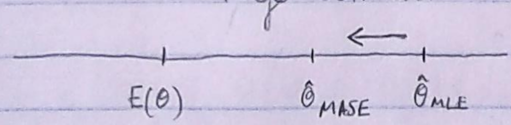
$$= \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right) E(\theta) + \left(\frac{n}{\alpha+\beta+n}\right) \cdot \hat{\theta}_{MLE}$$

$$\hat{\theta}_{MASE} = \rho\, E(\theta) + (1-\rho)\, \hat{\theta}_{MLE}$$

Note

$\frac{\alpha+\beta}{\alpha+\beta+n} = \rho$ | $\frac{n}{\alpha+\beta+n} = 1-\rho$

"Shrinkage estimator" – ρ acts as shrinkage proportion



$\rho = \frac{\alpha+\beta}{n+\alpha+\beta}$

$\alpha+\beta \gg n$ "strong" prior, small sample size

$\alpha+\beta \ll n$ "weak" prior, large sample size