Lecture 6: 2/22/21

$\hat{\theta}_{MMAE} := \text{Med}[\theta \mid X] = a$ such that $\int_{-\infty}^{a} P(\theta|x)\,d\theta = \frac{1}{2}$

Minimum Mean Absolute Error i.e. $\hat{\theta}_{MMAE} = \text{arg min}\left\{ E\left[|\theta - \hat{\theta}| \mid X\right]\right\}$

Using our model: iid Bern(θ) and data $x = \langle 0,1,1 \rangle$, we can compute the MMAE Bayesian point estimate:

$$\int_0^a 12\theta^2(1-\theta)\,d\theta = 12\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_0^a = 12\left(\frac{a^3}{3} - \frac{a^4}{4}\right) \overset{set}{=} \frac{1}{2}$$

$$-\frac{1}{4}a^4 + \frac{1}{3}a^3 + 0a^2 + 0a - \frac{1}{24} = 0 \Rightarrow a \approx 0.614$$

This is a "quartic equation" and has a formulaic solution that you can look up. The answer is: $a \approx 0.614$

These are the three Bayesian point estimates we will use for the rest of the class i.e. $\hat{\theta}_{MMSE}$, $\hat{\theta}_{MMAE}$, $\hat{\theta}_{MAP}$

The data $x = \langle 0,1,1 \rangle$ was a specific case. We will now solve this generally for any dataset $x = \langle x_1, \ldots, x_n \rangle$. Also using Laplace's prior of indifference, $\theta \sim U(0,1)$

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} = \frac{P(X|\theta)\cancel{P(\theta)}^{1}}{\int_0^1 P(X|\theta)\cancel{P(\theta)}^{1}d\theta} = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\int_0^1 \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}d\theta}$$

This integral in the denominator is a special integral and is known as the "beta function". The beta function has no closed form solution but can be calculated to arbitrary precision using a scientific calculator.

$$B(\alpha, \beta) := \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$$

$$B(\alpha,\beta) = \frac{1}{B(\sum x_i + 1, n - \sum x_i + 1)}\theta^{\sum x_i + 1 - 1}(1-\theta)^{n - \sum x_i + 1 - 1} = \text{Beta}\left(\sum x_i + 1, n - \sum x_i + 1\right)$$

We just derived that the posterior for the iid bernoulli likelihood is a beta distribution. Let's go back to probability class and examine the beta distribution...

$$Y \sim \text{Beta}(\alpha, \beta) \overset{PDF}{:=} \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1} = p(y)$$

$\text{Supp}[Y] = (0,1)$ $\quad \displaystyle\int_0^1 \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}\,dy = \frac{1}{\cancel{B(\alpha,\beta)}} \overbrace{\int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,dy}^{\cancel{B(\alpha,\beta)}} = 1$ ✓

$\alpha \in ?,\ \beta \in ?\quad \alpha > 0,\ \beta > 0$

$\alpha = 0,\ \beta = 1 \Rightarrow \int_0^1 \frac{1}{y}\,dy = \infty \Leftarrow$ doesn't work

$$E[Y] = \int_0^1 y\,p(y)\,dy = \int_0^1 y \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}\,dy = \frac{1}{B(\alpha,\beta)} \int_0^1 y^{\alpha+1-1}(1-y)^{\beta-1}\,dy = \underbrace{\frac{B(\alpha+1,\beta)}{B(\alpha,\beta)}}_{\text{to simplify this we need the gamma function.}}$$

$$\Gamma(\alpha) := \int_0^\infty t^{\alpha-1} e^{-t}\,dt,\quad \alpha > 0$$

Facts: ① $\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha)$, ② $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

$$\frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} \overset{②}{=} \frac{\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}}{B(\alpha,\beta)} \overset{①}{=} \cancel{\frac{\frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)}}{B(\alpha,\beta)}} \quad \frac{\alpha\,\frac{\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\,\Gamma(\alpha+\beta)}}{\cancel{B(\alpha,\beta)}} = \frac{\alpha}{\alpha+\beta}$$

$$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = B(\alpha,\beta)$$

$\text{Var}[Y] =$ on H.W.

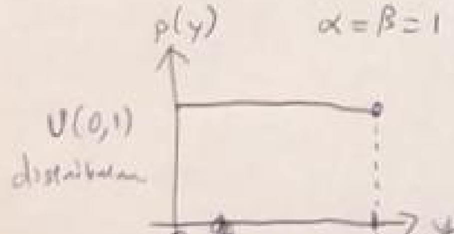
$$\text{Mode}[Y] = \underset{y\in(0,1)}{\arg\max}\left\{\frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}\right\} = \arg\max\left\{(\alpha-1)\ln(y) + (\beta-1)\ln(1-y)\right\}$$

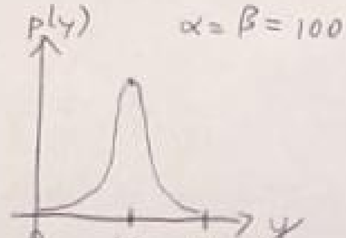
$$\overset{\text{deriv}}{\Rightarrow} \frac{\alpha-1}{y} - \frac{\beta-1}{1-y} \overset{\text{set}}{=} 0 \Rightarrow y_* = \frac{\alpha-1}{\alpha+\beta-2}$$

if we take the second derivative to check it's only negative if both α and β are greater than or equal to $\underline{1}$.

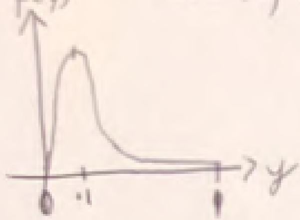
$\text{Med}[Y]$: has no closed form expression and thus must be done with a computer. We will denote the answer to this using notation from the R programming language: qbeta$(0.5, \alpha, \beta)$.

Let's take a look at the shapes of the Beta distribution:

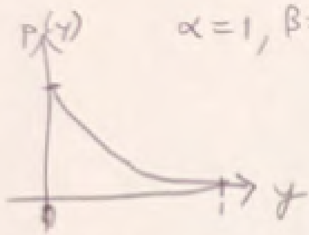




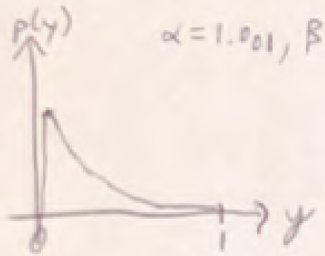
$\alpha = 10, \beta = 90$



$\alpha = 1, \beta = 4$

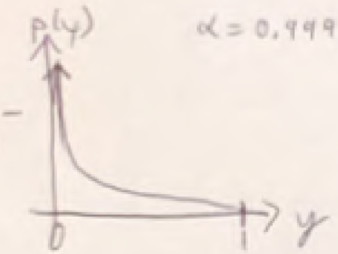


$\alpha = 1.001, \beta = 4$

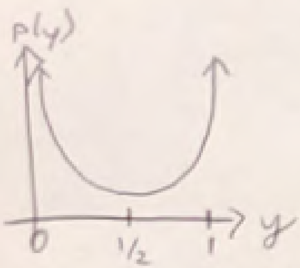


$\alpha = 0.999, \beta = 4$

no - mode

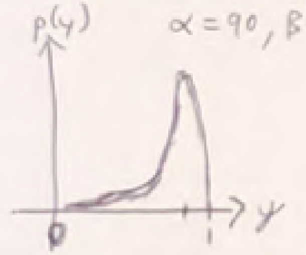


no - mode

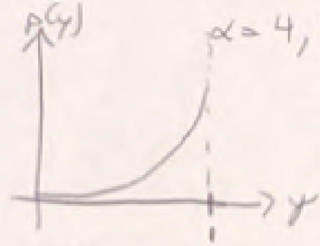


Arcsin

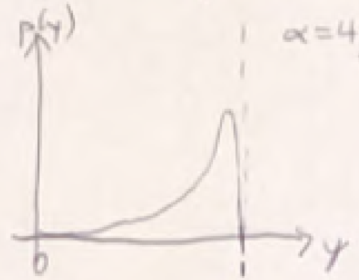
$\alpha = 90, \beta = 10$



$\alpha = 4, \beta = 1$



$\alpha = 4, \beta = 1.001$



$\alpha = 4, \beta = 0.999$

- no mode