11/2/2021. LECTURE 4.

Last time: $f(\theta) = \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$

Beta distribution

showed that $E(\theta) = \frac{\alpha}{\alpha+\beta}$. Similarly $E(\theta^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$

Thus, $Var(\theta) = E(\theta^2) - E^2(\theta) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2$

$$= \frac{\alpha}{\alpha+\beta}\left[\frac{(\alpha+1)(\alpha+\beta) - \alpha(\alpha+\beta+1)}{(\alpha+\beta)(\alpha+\beta+1)}\right] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

[Symmetric in α, β, $Var(1-\theta) = Var(\theta)$]

The Beta distribution is a very flexible distribution and α, β control the shape of the density function.

[See plot at the end].

Consider the posterior distribution when $\theta \sim Beta(\alpha,\beta)$, $X|\theta \sim Bin(n,\theta)$

$$f(\theta|x) \propto \mathbb{P}(X=x|\theta) f(\theta)$$

$$f(\theta|x) \propto \binom{n}{x} \theta^x (1-\theta)^{n-x} \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

wrt θ

$$\propto \theta^x(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$
$$= \theta^{(\alpha+x)-1}(1-\theta)^{(\beta+n-x)-1}$$

In general, $f(\theta|x) = c\, g(\theta)$ for some constant c not involving θ. As $\int_\theta f(\theta|x)\, d\theta = 1$ then

$$c = \left[\int_\theta g(\theta)\, d\theta\right]^{-1}$$

[NB. $f(\theta|x) \propto f(x|\theta) f(\theta) \propto g(\theta)$ when $g(\theta) = \theta^{(\alpha+x)-1}(1-\theta)^{(\beta+n-x)-1}$ by dropping the constant terms (in terms of θ) $\binom{n}{x}$ from $f(x|\theta)$ and $\frac{1}{B(\alpha,\beta)}$ from $f(\theta)$].

In many cases, this integral could be difficult to calculate. In this case,

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

so that $$f(\theta|x) = \frac{1}{B(\alpha+x, \beta+n-x)} \theta^{(\alpha+x)-1}(1-\theta)^{(\beta+n-x)-1}$$

i.e. $\theta|x \sim \text{Beta}(\alpha+x, \beta+n-x)$.

Effectively we revise, or update,

$\alpha \mapsto \alpha + x$ (add the number of successes)

$\beta \mapsto \beta + n - x$ (add the number of failures)

Note the tractability of this: the prior and the posterior are from the same family of distributions. This is an example of CONJUGACY.

Definition (Conjugacy).
A class π of prior distributions is said to form a CONJUGATE FAMILY with respect to a likelihood $f(x|\theta)$ if the posterior density is in the class π for all x whenever the prior density is in π.

We've shown that with respect to the Binomial likelihood, the Beta distribution is a conjugate prior.

We'll now look at the effect of the prior on the posterior.

Example
We return to the motivating example and consider two Beta-Binomials

Set up so that I observe the number of successes in the same number of trials.

① Tossing coins, parameter θ_c

Prior $\theta_c \sim \text{Beta}(\alpha_c, \beta_c)$

Likelihood $X_c | \theta_c \sim \text{Bin}(n, \theta_c)$

Posterior $\theta_c | X_c = x \sim \text{Beta}(\alpha_c + x, \beta_c + n - x)$.

② Tossing drawing pins, parameter θ_p

Prior $\theta_p \sim \text{Beta}(\alpha_p, \beta_p)$

Likelihood $X_p | \theta_p \sim \text{Bin}(n, \theta_p)$

Posterior $\theta_p | X_p = x \sim \text{Beta}(\alpha_p + x, \beta_p + n - x)$

I have more prior knowledge about θ_c than θ_p. In each case, suppose that I suspect, a priori, that

$$E(\theta_c) = \frac{1}{2} = E(\theta_p)$$

[$\theta \sim \text{Beta}(\alpha, \beta)$ if $E(\theta) = \frac{1}{2}$ then $\frac{\alpha}{\alpha+\beta} = \frac{1}{2}$ i.e. $\underline{\underline{\alpha = \beta}}$]

For θ_c I expect a smaller variance than θ_p. I assess

$\theta_c \sim \text{Beta}(20, 20)$

[$\text{Var}(\theta_c) = 0.006$]

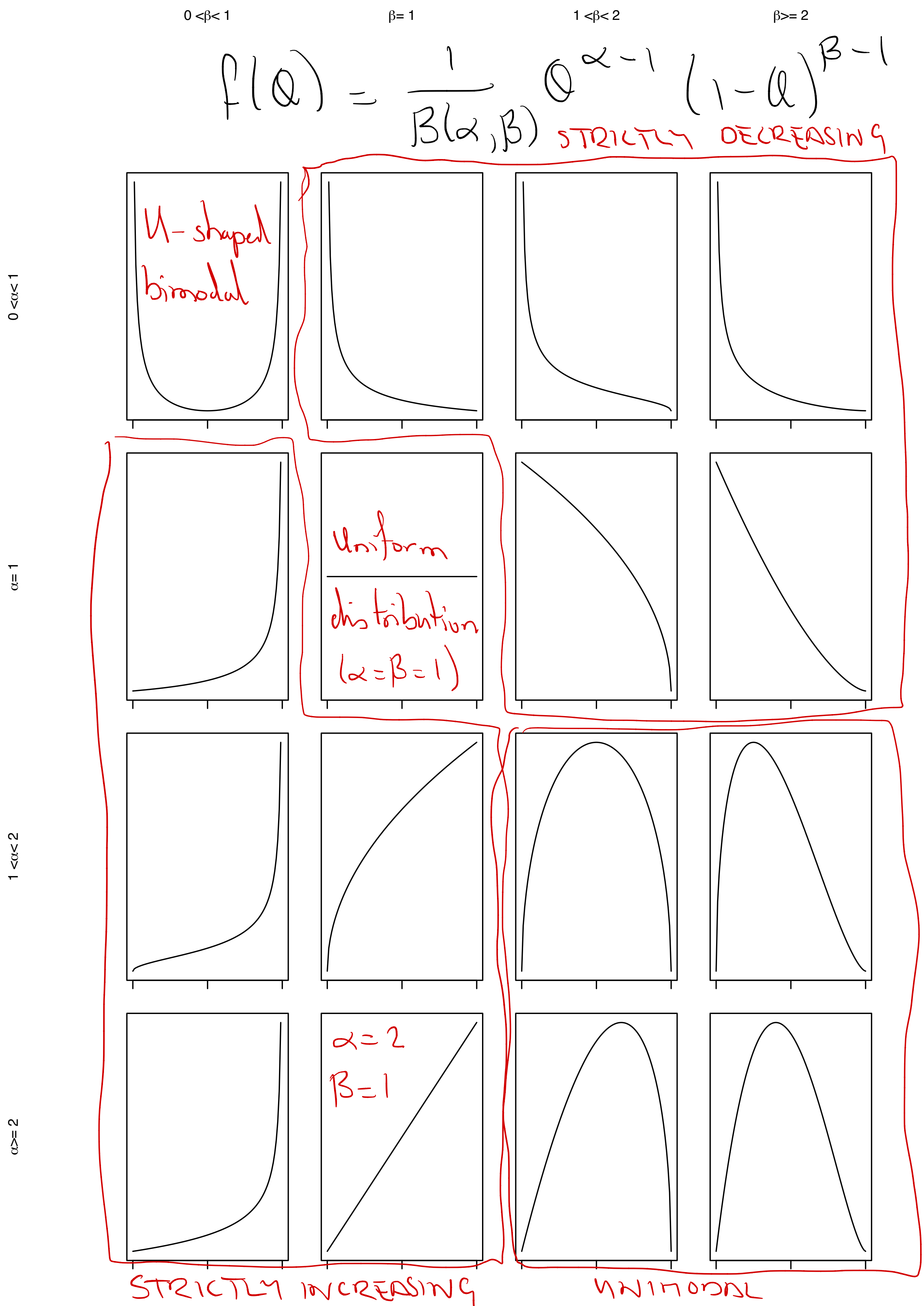
$\theta_p \sim \text{Beta}(2, 2)$

[$\text{Var}(\theta_p) = 0.05$]

[$\theta \sim \text{Beta}(\alpha, \beta)$, $\text{Var}(\theta) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

If $\alpha = \beta$ $\text{Var}(\theta) = \dfrac{1}{4(2\beta+1)}$

Larger $\beta \Rightarrow$ smaller variance].



$f(\theta) = \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$
0 <β< 1
β= 1
1 <β< 2
β>= 2
0 <α< 1
α= 1
1 <α< 2
α>= 2
STRICTLY DECREASING
U-shaped bimodal
Uniform distribution (α=β=1)
α=2 β=1
STRICTLY INCREASING
UNIMODAL