Summary of Conjugate Priors

Assume $x_i \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$ are i.i.d.

Prior for μ: $\mu \mid \sigma^2 \sim N(\beta, \tau^2)$ where τ^2 possibly function of σ^2

Posterior for μ: $$\mu \mid x_1, \dots, x_n, \sigma^2 \sim N\left(\frac{\sigma^2 \beta/n + \tau^2 \bar{x}}{\sigma^2/n + \tau^2}, \frac{\sigma^2/n \, \tau^2}{\sigma^2/n + \tau^2} \right)$$

Prior for σ^2: $$\left. \frac{S_0 + n_0(\mu - \beta)^2}{\sigma^2} \right| \mu \sim \chi^2_k$$ (n_0 possibly 0)

Posterior for σ^2: $$\left. \frac{\sum (x_i - \mu)^2 + S_0 + n_0(\mu - \beta)^2}{\sigma^2} \right| x_1, \dots, x_n, \mu \sim \chi^2_{k+n}$$

Posterior for μ and σ^2: $\quad \mu \mid \sigma^2 \sim N\left(\beta, \frac{\sigma^2}{n_0}\right)$, $\quad \frac{S_0}{\sigma^2} \sim \chi^2_{k_0}$

$$\Longrightarrow \quad \frac{\mu - \beta}{\sqrt{S_0 / n_0 k_0}} \sim t_{k_0}$$

$$\frac{S_0 + n_0(\mu - \beta)^2}{\sigma^2} \,\Big|\, \mu \sim \chi^2_{k_0 + 1}$$

Posterior for μ and σ^2: $\quad \frac{S'}{\sigma^2} \sim \chi^2_k \qquad \frac{\mu - \beta'}{\sqrt{S' / n' k}} \sim t_k$

$$k = n + k_0$$

$$\beta' = \frac{n_0 \beta + n \bar{x}}{n + n_0} \qquad S = \sum (x_i - \bar{x})^2$$

$$n' = n_0 + n$$

$$S' = S + S_0 + n_0 \beta^2 + n \bar{x}^2 - n' \beta'^2$$

The Beta density:

Suppose $x_i \mid p \sim \text{Ber}(p)$ are i.i.d.

Then if $S_n = X_1 + X_2 + \dots + X_n$, S_n is Binomial conditioned on p.

$$P(S_n = k \mid p) = \binom{n}{k} p^k (1-p)^{n-k}$$

e.g. Tossing unknown coin n times

With p unknown, what is an appropriate prior on p?

Consider Bayes rule.

$$f(p \mid S_n = k) \propto P(S_n = k \mid p) f(p) \propto p^k (1-p)^{n-k} f(p)$$

So if $f(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$

Then $f(p \mid S_n = k) \propto p^{(k+\alpha)-1}(1-p)^{(n-k+\beta)-1}$

A density of the form

$$f(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$$ would be conjugate.

It turns out $\displaystyle\int_0^1 p^{\alpha-1}(1-p)^{\beta-1}\,dp = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ $\quad \alpha,\beta > 0$

So $$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, p^{\alpha-1}(1-p)^{\beta-1} \quad 0 \leq p \leq 1 \quad \alpha,\beta > 0$$

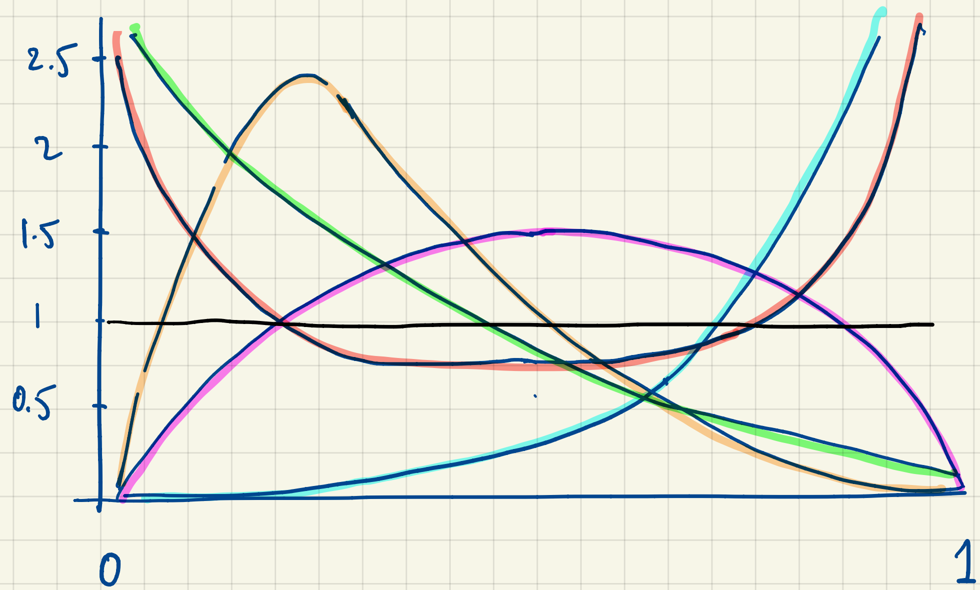
is a valid prob. density function.

This is called the Beta density, with parameter α and β

$$E[p] = \frac{\alpha}{\alpha+\beta} \qquad Var(p) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

The Beta density is appropriate for modeling prior on probability

By changing α and β, we can achieve different shapes.



When $\alpha = \beta = 1$, it's uniform in $[0,1]$

$$\alpha = \beta = 1 \Rightarrow f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} p^{0}(1-p)^{0} = \frac{1!}{0!\,0!} 1 \cdot 1 = 1$$

<u>Example:</u> Given a bin with red and black balls, the prob. of red is p, but p is unknown.

We repeatedly pick balls with replacement.

Assume Uniform prior on p.

What is the posterior density of p after observing $\alpha-1$ reds and $\beta-1$ blacks? $[n = \alpha-1+\beta-1 = \alpha+\beta-2]$

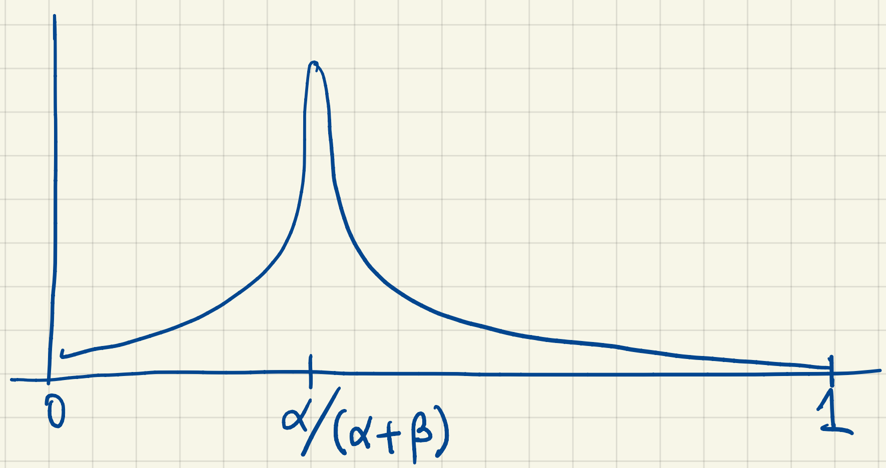
$$f(p \mid \alpha-1 \text{ red}, \beta-1 \text{ black}) \propto P(\alpha-1 \text{ red}, \beta-1 \text{ black} \mid p)\, f(p)$$

$$= \binom{\alpha+\beta-2}{\alpha-1} p^{\alpha-1}(1-p)^{\beta-1} \cdot 1$$

Poseterior for p is $\text{Beta}(\alpha, \beta)$

observation		posterior
$\alpha=1$	$\beta=1$	$f(p)=1$
$\alpha=2$	$\beta=1$	$f(p)=2p$
$\alpha=2$	$\beta=2$	$f(p)=6p(1-p)$
$\alpha=3$	$\beta=1$	$f(p)=3p^2$
$\alpha=3$	$\beta=2$	$f(p)=12p^2(1-p)$
$\alpha=3$	$\beta=3$	$f(p)=30p^2(1-p)^2$

When both α and β are large, $f(p)$ will look like:



Laplace's Rule of Succession

What is the prob. that the sun will rise tomorrow ?

The sun rise problem.

How to model this ?

$$P(X_{n+1} = 1 \mid S_n = k)$$

$$P(A) = \int P(A \mid x) f(x) dx$$

$$P(A \mid B) = \int P(A \mid B, x) f(x \mid B) dx$$

$$\int_0^1 P(X_{n+1} = 1 \mid p, S_n = k) f(p \mid S_n = k) dp = \int_0^1 P(X_{n+1} = 1 \mid p) f(p \mid S_n = k) dp$$

$$= \int_0^1 p f(p \mid S_n = k) dp = E[p \mid S_n = k]$$

Assume we don't know much, so $p \sim \text{Unif}(0,1)$ $\quad f(p) = 1$

$$f(p \mid S_{n=k}) \propto p^k (1-p)^{n-k}$$

So $$f(p \mid S_n = k) = \frac{\Gamma(k+1+n-k+1)}{\Gamma(k+1)\Gamma(n-k+1)} p^{(k+1)-1} (1-p)^{(n-k+1)-1}$$

$$\int_0^1 p \frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)} p^{(k+1)-1} (1-p)^{(n-k+1)-1} dp = E[p \mid S_n = k]$$

$$P(X_{n+1} = 1 \mid S_n = k) = \frac{k+1}{n+2} \quad (\text{if } k = n \to \infty, \approx 1)$$

Interpretation: Since we have k successes, we think $p = \frac{k}{n}$. In fact this is the maximum likelihood probability, i.e. $p = \frac{k}{n}$ maximizes $P(S_n = k \mid p)$

- But what if $k=n$, then $p=1$ is not a satisfying answer. Typically, we add a pseudo count to avoid zero.

$$\left.\begin{aligned} k &\leftarrow k + a \\ n-k &\leftarrow n-k+b \\ n &\leftarrow n+a+b \end{aligned}\right| \Rightarrow p = \frac{k+a}{n+a+b}$$

- The pseudocounts reflect some "prior" belief. For instance if we believe $p \approx 0.5$, we can make $a=b$ large. Then k and n will not be able to skew this belief.
- Laplace's approach is somewhat equivalent to a pseudo count of $a=b=1$, but it's conceptually cleaner.

Binomial $| p$

prior

$f(p)$

$\propto p^{\alpha-1}(1-p)^{\beta-1}$

observe

k success

and $n-k$ failures

$f(p \mid S_n = k)$

$\propto p^{(k+\alpha)-1}(1-p)^{(n-k+\beta)-1}$

Beta density