1. Let $X = \{x_1, x_2, \dots, x_n\}$

$$p(x_i|\theta) = \begin{cases} \theta x_i^{\theta-1} & ; \ 0 < x < 1 \\ 0 & ; \ \text{elsewhere} \end{cases}$$

$$\theta \sim \Gamma(\alpha, \beta) \Rightarrow p(\theta) = \frac{\beta^\alpha \theta^{\alpha-1} e^{-\beta\theta}}{\Gamma(\alpha)}$$

$$p(X|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$$

$$= \theta^n \left(\prod_{i=1}^{N} x_i\right)^{\theta-1}$$

$$p(\theta|X) \propto p(X|\theta)\, p(\theta)$$

$$= \left(\prod_{i=1}^{N} x_i\right)^{\theta} \theta^{\alpha+n-1} e^{-\beta\theta}$$

[Altering constant factors as per convenience]

Let $\prod_{i=1}^{n} x_i = X_p \Rightarrow X_p^\theta = e^{\ln(X_p)\theta}$

$$\Rightarrow p(\theta|X) \propto \theta^{\alpha+n-1} \cdot \exp(-(\beta - \ln(X_p))\theta)$$

This is clearly a gamma distribution, so we can directly write the $p(X)$ marginal as this is a known integral.

$$p(\theta|X) = \frac{(\beta - \ln(X_p))^{\alpha+n} \cdot \theta^{\alpha+n-1} \cdot \exp((\ln X_p - \beta)\theta)}{\Gamma(\alpha+n)}$$

or in other words

$$\theta|X \sim \Gamma(\alpha+n, \beta - \ln X_p)$$

There doesn't seem to be any closed form expression for $\eta(x)$, so we leave it like so.

2. Posterior $\propto$ Prior $\times$ Likelihood

Given k, α, β,

$$p(\theta|k,\alpha,\beta) \propto (1-\theta)^{k-1}\theta \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$= \theta^\alpha (1-\theta)^{\beta+k-2}$$

$$\Rightarrow p(\theta|k,\alpha,\beta) = \eta(k,\alpha,\beta)\, \theta^\alpha (1-\theta)^{\beta+k-2}$$

It is easy to see that

$$\eta(\alpha,\beta,k) = \int_0^1 \theta^{\alpha} (1-\theta)^{\beta+k-2} \, d\theta$$

which is simply the Beta function integral with parameters $\alpha' = \alpha+1$, $\beta' = \beta+k-1$

$$\Rightarrow \eta(\alpha,\beta,k) = \frac{\Gamma(\alpha+\beta+k)}{\Gamma(\alpha+1)\Gamma(\beta+k-1)}$$

$$\therefore p(\theta|k,\alpha,\beta) = \frac{\Gamma(\alpha+\beta+k)}{\Gamma(\alpha+1)\Gamma(\beta+k-1)} \theta^{\alpha} (1-\theta)^{\beta+k-2}$$

Which is a Beta distribution, hence proved.

3. $K = \{k_1, k_2, \dots, k_n\}$

$$p(K|\theta) = \prod_{i=1}^{n} (1-\theta)^{k_i-1} \theta$$

$$= (1-\theta)^{\sum_{i=1}^{n} k_i - n} \theta^{n}$$

$$p(\theta|K,\alpha,\beta) = \frac{p(K|\theta,\alpha,\beta)\, p(\theta|\alpha,\beta)}{\int_0^1 p(K|\theta,\alpha,\beta)\, p(\theta|\alpha,\beta)\, d\theta}$$

$$= \frac{\Gamma(\alpha+\beta+\sum_{i=1}^{n} k_i)}{\Gamma(\alpha+n)\Gamma(\beta+\sum_{i=1}^{n} k_i - n)} \theta^{\alpha+n-1} (1-\theta)^{\beta+\sum_{i=1}^{n} k_i - n - 1}$$

This is the posterior, which still follows the Beta distribution: $\mathrm{Beta}(\alpha+n, \beta+\sum_{i=1}^{n} k_i - n)$

Accordingly, from the properties of a Beta distribution we have:

$$\hat{\theta}_{mean} = \frac{\alpha+n}{\alpha+\beta+\sum_{i=1}^{n} k_i}$$

$$\hat{\theta}_{mode} = \frac{\alpha+n-1}{\alpha+\beta+\sum_{i=1}^{n} k_i - 2}$$

α and β, known as the shape parameters of a Beta distribution, skew the likelihood of θ towards 1 or 0. $\alpha \gg \beta$ makes the distribution prefer numbers closer to 1 and $\beta \gg \alpha$ makes the distribution prefer numbers closer to 0. This can be seen in that α being the exponent of θ means small values of θ are very improbable, and a similar argument can be made for $(1-\theta)$ and β.
$\hat{\theta}_{mean}$ & $\hat{\theta}_{mode}$ can also tell us about the distribution of θ.

$\alpha \gg \beta \Rightarrow \hat{\theta}_{mean} \approx 1$, $\hat{\theta}_{mode} \approx 1$

$\alpha \ll \beta \Rightarrow \theta_{mean} \approx \alpha/\beta$, $\hat{\theta}_{mode} \approx \alpha/\beta$, and α/β is clearly a very small quantity.

As $n \longrightarrow \infty$, the effect on the distribution of θ depends upon the values of k_i.

$$k_i > 1 \quad \forall\, i \quad \Rightarrow \quad \sum_{i=1}^{n} k_i > n$$

This means $\sum_{i=1}^{n} k_i > n \gg \alpha, \beta, 1$

$$\Rightarrow \hat{\theta}_{mean} \approx \hat{\theta}_{mode} \approx \frac{n}{\sum_{i=1}^{n} k_i}.$$

If the values of k_i are small, we see that results in $\alpha > \beta$, which means the distribution of θ favours larger values. This makes intuitive sense since if θ is close to 1, a geometric progression caused by bernoulli trials will end early since success is more probable than failure.

If values of k_i are much bigger, then $\beta > \alpha$ and small values of θ are likelier to be drawn from the prior distribution. This also makes intuitive sense since if θ is close to 0, it is easy to see a large number of subsequent bernoulli trials are needed to find success.

4. $Y = Y_{1,\dots,N}$ s.t. $Y_i \sim \mathcal{N}(\beta x_i, \sigma^2)$

a. $\beta \sim \mathcal{N}(\beta_0, \tau^2)$, we find the posterior distribution $p(\beta|Y)$

$$p(\beta|Y) \propto p(Y|\beta) \cdot p(\beta)$$

$$\propto \exp\left(\frac{(\beta-\beta_0)^2}{2\tau^2}\right) \cdot \left(\prod_{i=1}^{N} \exp\left(\frac{(Y_i - \beta x_i)^2}{2\sigma^2}\right)\right)$$

$$\propto \exp\left(\frac{(\beta-\beta_0)^2}{2\tau^2}\right) \cdot \exp\left(\frac{1}{2}\sum_{i=1}^{N} \frac{(\beta - Y_i/x_i)^2}{(\sigma^2/x_i^2)}\right)$$

Keeping $\delta_i = Y_i/x_i$ and $\sigma_i^2 = \sigma^2/x_i^2$, we analyse the following expression

$$\sum_{i=1}^{N} \frac{(\beta-\delta_i)^2}{\sigma_i^2} = \beta^2 \sum_{i=1}^{N} \frac{1}{\sigma_i^2} - 2\beta \sum_{i=1}^{N} \frac{\delta_i}{\sigma_i^2} + K$$, where K is some constant we can ignore because only proportionality matters

$$= \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left(\beta^2 - 2\beta \frac{\sum_{i=1}^{N} \delta_i/\sigma_i^2}{\sum_{i=1}^{N} 1/\sigma_i^2} + K\right)$$

$$= \sum \frac{1}{\sigma_i^2} \left(\beta - \frac{\sum_{i=1}^{N} \delta_i/\sigma_i^2}{\sum_{i=1}^{N} 1/\sigma_i^2}\right)^2 + K$$

$\Rightarrow$ The product of gaussians is also a gaussian $\mathcal{N}(\mu', \sigma'^2)$ where

$$\frac{1}{\sigma'^2} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \quad \text{and} \quad \mu' = \frac{\sum_{i=1}^{N} \delta_i/\sigma_i^2}{\sum_{i=1}^{N} 1/\sigma_i^2}$$

$$= \frac{\sum_{i=1}^{N} x_i^2}{\sigma^2}, \qquad = \frac{\sum_{i=1}^{N} x_i Y_i/\sigma^2}{\sum_{i=1}^{N} x_i^2/\sigma^2} = \frac{\sum_{i=1}^{N} x_i Y_i}{\sum_{i=1}^{N} x_i^2} \qquad (i)$$

Substituting this back, we get

$$p(\beta | Y) \propto \exp\left(\frac{(\beta - \beta_0)^2}{2\tau^2}\right) \cdot \exp\left(\frac{(\beta - \mu')^2}{2\sigma'^2}\right)$$

$$= \exp\left(\frac{1}{2}\beta^2\left(\frac{1}{\tau^2} + \frac{1}{\sigma'^2}\right) - \beta\left(\frac{\beta_0}{\tau^2} + \frac{\mu'}{\sigma'^2}\right) + K\right)$$

$$= \exp\left(\frac{1}{2}\left(\frac{1}{\tau^2} + \frac{1}{\sigma'^2}\right)\left(\beta^2 - 2\beta\left(\frac{\beta_0/\tau^2 + \mu'/\sigma'^2}{1/\tau^2 + 1/\sigma'^2}\right) + K\right)\right)$$

$$= \exp\left(\frac{(\beta - \mu'')^2}{2\sigma''^2}\right) \qquad \text{(ii)}$$

This clearly implies that $p(\beta|Y)$ is Normally distributed $\mathcal{N}(\mu'', \sigma''^2)$, where

$$\mu'' = \frac{\beta^0/\tau^2 + \sum_{i=1}^{N} \frac{x_i Y_i}{\sigma^2}}{1/\tau^2 + \sum_{i=1}^{N} x_i^2/\sigma^2}$$

$$\frac{1}{\sigma''^2} = \frac{1}{\tau^2} + \sum_{i=1}^{N} \frac{x_i^2}{\sigma^2}$$

(b) Prior mean $= \beta_0$

$$\hat{\beta}_{MLE} = \underset{\beta}{\operatorname{argmax}}\left(\prod_{i=1}^{N} p(Y_i | \beta)\right)$$

$$= \underset{\beta}{\operatorname{argmax}}\left(\exp\left(\frac{Y_i - x_i\beta}{2\sigma^2}\right)\right)$$

$$= \mu'$$ [This expression is evaluated in (i)]

Posterior mean $= \mu''$

$$= \frac{\beta_0/\tau^2 + \mu'/\sigma'^2}{1/\tau^2 + 1/\sigma'^2}$$ [From (ii)]

Which is clearly a weighted mean of β_0 and μ'.

c. As $\tau^2 \longrightarrow \infty$, we can see that

$\mu'' = \mu'$
$\sigma''^2 = \sigma'^2$

This intuitively makes sense because $\tau^2 \longrightarrow \infty$ means the prior is uniformly distributed over the entire real line, and hence offers no additional information when we evaluate the posterior from the likelihood. The prior now contains no information about the structure of the posterior, hence the posterior now depends entirely on the likelihood, i.e. on the data.

d. For a future observation we need to evaluate $p(Y_{n+1} \mid Y, x_{n+1})$. We express this as a marginal from the joint $p(Y_{n+1}, \beta \mid Y, x_{n+1})$, where Y_{n+1} is the new observation and $Y = Y_{1,\dots N}$.

$$
\begin{aligned}
p(Y_{n+1} \mid Y, x_{n+1}) &= \int_{\beta \in \mathbb{R}} p(Y_{n+1}, \beta \mid Y, x_{n+1})\, d\beta \\
&= \int_{\beta \in \mathbb{R}} p(Y_{n+1} \mid \beta, Y, x_{n+1})\, p(\beta \mid Y, x_{n+1})\, d\beta \\
&= \int_{\beta \in \mathbb{R}} \mathcal{N}(Y_{n+1}; \beta x_{n+1}, \sigma^2)\, \mathcal{N}(\beta; \mu'', \sigma''^2)\, d\beta \\
&= \int_{\beta \in \mathbb{R}} \exp\left(\frac{(Y_{n+1} - \beta x_{n+1})^2}{2\sigma^2}\right) \exp\left(\frac{(\beta - \mu'')^2}{2\sigma''^2}\right) d\beta
\end{aligned}
$$

We don't exactly need the distribution of $Y_{n+1} \mid Y, x_{n+1}$, we only need the MAP estimate. To evaluate that we observe that both products are concave down which means differentiating wrt Y_{n+1} and equating to zero would give us the prediction for Y_{n+1}. So we have,

$$
\frac{d}{dY_{n+1}} \int_{-\infty}^{\infty} \exp\left(\frac{(Y_{n+1} - \beta x_{n+1})^2}{2\sigma^2}\right) \exp\left(\frac{(\beta - \mu'')^2}{2\sigma''^2}\right) d\beta = 0
$$

$$
\Rightarrow \int_{-\infty}^{\infty} \frac{(Y_{n+1} - \beta x_{n+1})}{\sigma^2} \exp\left(\frac{(Y_{n+1} - \beta x_{n+1})^2}{2\sigma^2}\right) \exp\left(\frac{(\beta - \mu'')^2}{2\sigma''^2}\right) d\beta = 0
$$

This is intractable to the best of my abilities so we use the <u>plug-in predictive</u> approximation

$$
\begin{aligned}
p(Y_{n+1} \mid Y, x_{n+1}) &\approx p(Y_{n+1} \mid \hat{\beta}_{MAP}, x_{n+1}) \\
&\propto \exp\left(\frac{Y_{n+1} - \hat{\beta}_{MAP} x_{n+1}}{2\sigma^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \text{The MAP estimate of } Y_{n+1} &= \hat{\beta}_{MAP}\, x_{n+1} \\
&= \mu''\, x_{n+1} \\
&= \frac{\beta^0/\tau^2 + \sum_{i=1}^{N} \frac{x_i Y_i}{\sigma^2}}{1/\tau^2 + \sum_{i=1}^{N} x_i^2/\sigma^2} \cdot x_{n+1}
\end{aligned}
$$