Tues morning, 23rd Sept 2025

Tutorial 2: Playing with probabilities

1. Conjugate prior: Beta

Ex 16.2 from *Mathematical Statistics with Applications* by Wackerly, Mendenhall and Scheafer.

In lecture you found that prior $\text{beta}(\alpha, \beta)$ with a binomial likelihood, yields a beta posterior with new parameters $\alpha^* = \alpha + k, \beta^* = n - k + \beta$

Suppose you're an epidemiologist studying a rare disease with probability p.

You know the disease is rare (maybe $< p > \approx 0.25$), and you want to include the rarity of p in the statistical analysis that you're making.

Since you're psyched by Bayesian methods, you're gointg to consider p as a random variable with prior given by a Beta distribution.

1a) (Warm-up): What's the mean and variance of the beta distribtuion?

Step 1: Calculate the mean

$$\mathbb{E}[x] = \int_0^1 x \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1} = \frac{1}{B(\alpha,\beta)} B(\alpha+1,\, \beta)$$

Where we identify the integral defining the beta function with an updated α.

Now use the definition of the Beta function: $B(\alpha, \beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ to simplify further

$$\mathbb{E}[x] = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)}$$

Finally, recall $\Gamma(z+1) = z\Gamma(z)$:

$$\mathbb{E}[x] = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} = \boxed{\frac{\alpha}{\alpha+\beta}}$$

Step 2: Calculate the variance

Use the theorem $\text{Var}(x) = \mathbb{E}[x^2] - \mathbb{E}[x]^2$

$$\mathbb{E}[x^2] = \int_0^1 x^2 \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{1}{B(\alpha,\beta)} B(\alpha+2,\, \beta)$$

Again plugging in the gammas:

$$\mathbb{E}[x^2] = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+2+\beta)} = \frac{(\alpha+1)\alpha}{(\alpha+1+\beta)(\alpha+\beta)}$$

Finally solve for the variance

$$\begin{aligned}\text{Var}(x) &= \mathbb{E}[x^2] - \mathbb{E}[x]^2 \\ &= \frac{\alpha(\alpha+1)}{(\alpha+1+\beta)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\ &= \frac{\alpha(\alpha+1)(\alpha+\beta) - \alpha^2(\alpha+\beta+1)}{(\alpha+1+\beta)(\alpha+\beta)^2}\end{aligned}$$

Let's simplify the numerator:

$$\text{num} = \alpha(\alpha+1)(\alpha+\beta) - \alpha^2(\alpha+\beta+1) = \alpha(\alpha^2+\alpha+\alpha\beta+\beta) - \alpha^3 - \alpha^2\beta - \alpha^2 = \alpha\beta$$

$$\text{Var}(x) = \boxed{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+1+\beta)}}$$

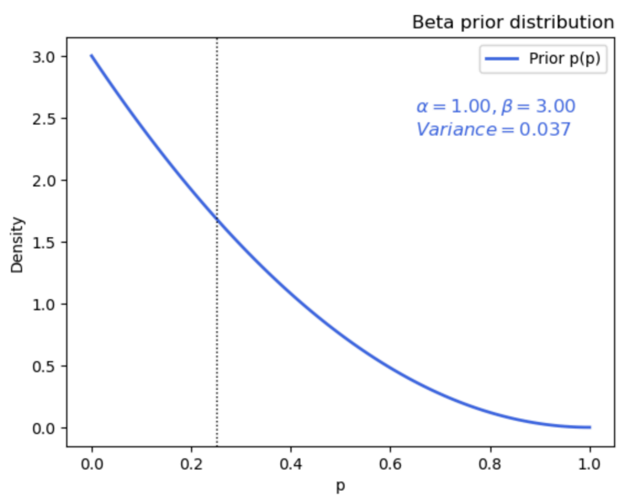
1b) For your epidemiology analayis, what α and β might you pick for the prior for $p(p)$?

Soln: $\mathbb{E}[x] = \dfrac{\alpha}{\alpha+\beta}$, so $\alpha = 1, \beta = 3$ would give the desired mean of 0.25.

(Nb: any soln that has $\alpha/(\alpha+\beta) = 1/4$ is an equally valid soln.)

1c) Plot this prior distribution that you're choosing

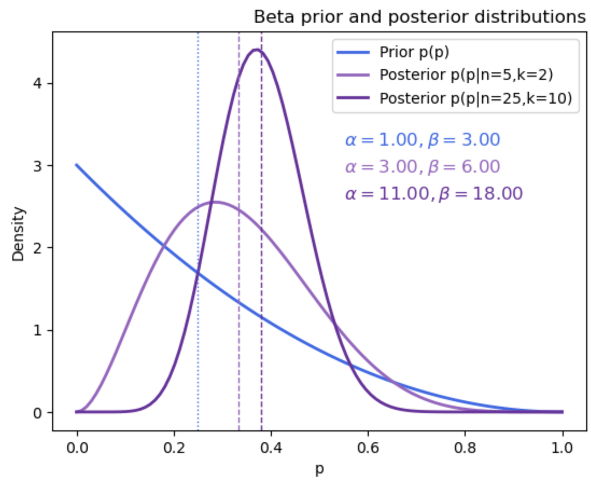
Soln: Use whichever plotting software you're most familiar with, here in python.

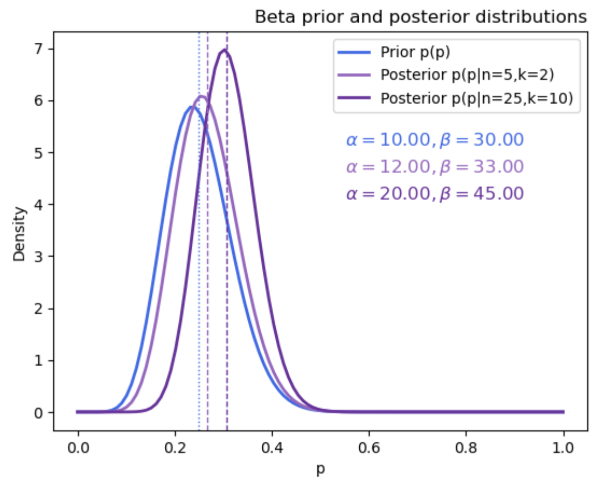


1d) Compare the prior and posterior distributions of the Bernoulli parameter p (proportion of responders to the new therapy) if we chose values for α and β given by the hypothetical data below:

a) $\alpha = 1, \beta = 3, n = 5, k = 2$
b) $\alpha = 1, \beta = 3, n = 25, k = 10$
c) $\alpha = 10, \beta = 30, n = 5, k = 2$
d) $\alpha = 10, \beta = 30, n = 25, k = 10$

Soln: Plotting in python the distributions and parameters.





Note — the mean of the binomial distribtuion is np, so all of these "data" in (a) — (d) have an expected <p> = .4 from the likelihood only.

In the left plot, we see the impact of increasing the size of the dataset. As we gather more responders in the experimental trial, the distribution becomes more peaked, and close to the 0.4 mean preferred by the data.

In the right plot, we see the impact of increasing the concentration of the prior distribution.
- When the prior is more peaked, the posterior becomes harder to update, none of the updated data are close the the MLE estimate for p of 0.4.

2. Conguate priors: Gamma and Poisson

2a) Yesterday in the tutorial, we encountered the Poisson distribution, which is characterized by a length parameter, λ:

$$P(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Q for you: of the distributions that we've encountered so far… which would be the a good candidate for modelling the rate parameter, λ ?

The rate parameter of the Poisson is some $\lambda > 0$.
The **Gamma distribution** models a postive variable (in your formula sheet we called it $x > 0$)

$$P(\lambda;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda/\beta}$$

Note, as the exponential and chi2 are also special cases of the Gamma distribution, these would also be valid choices for the prior on λ , but here we work through the example with the full generality of the Gamma prior.

2b) Consider n samples from the Poisson probability with outcomes $k_1, k_2, \ldots, k_n$.
For the prior you proposed in part (a), solve for the posterior probability.

Hint: if you define $k = \sum_{i=1}^{n} k_i$ to make the maths a bit more succinct

$$P(\lambda \,|\, k_1, k_2, \ldots, k_n) = \frac{P(k_1, k_2, \ldots, k_n \,|\, \lambda) P(\lambda)}{P(k)}$$

The likelihood is the sum of n independent trials, so using the independence property:

$$P(k_1, k_2, \ldots, k_n \,|\, \lambda) = \prod_i P(k_i \,|\, \lambda) = \prod_i \frac{\lambda_i^k e^{-\lambda}}{k_i!} = \frac{\lambda^{\sum_i k_i} e^{-n\lambda}}{\prod_i k_i!} = \frac{\lambda^k e^{-n\lambda}}{\prod_i k_1!}$$

I'll simplify the notation a bit in the following to solve for $P(\lambda \,|\, k)$ instead of $P(\lambda \,|\, k_1 \ldots k_n)$ since in the end the sum of the ks will be the only thing the posterior depends on is this k.

Let's follow the solution strategy from class.

Numerator:

$$
\begin{aligned}
P(k\,|\,\lambda)\ P(\lambda;\alpha,\beta) &= P(k_1,k_2,\ldots,k_n\,|\,\lambda)\ P(\lambda;\alpha,\beta)\\
&= \frac{\lambda^k e^{-n\lambda}}{\prod_i k_1!}\frac{1}{\beta^\alpha\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda/\beta}\\
&= \frac{1}{\prod_i k_1!}\frac{1}{\beta^\alpha\Gamma(\alpha)}\lambda^{\alpha+k-1}e^{-(n+1/\beta)\lambda}
\end{aligned}
$$

Solving for the effective β^* in the exponent:

$$
n + 1/\beta = \frac{n\beta+1}{\beta} \qquad \Longrightarrow \qquad \beta^* = \frac{\beta}{n\beta+1}
$$

Denominator: integrating over λ to get the evidence

$$
\begin{aligned}
P(k) &= \int_0^\infty d\lambda P(k\,|\,\lambda)P(\lambda;\alpha,\beta)\\
&= \int_0^\infty d\lambda \frac{1}{\prod_i k_1!}\frac{1}{\beta^\alpha\Gamma(\alpha)}\lambda^{\alpha+k-1}e^{-(n+1/\beta)\lambda}\\
&= \frac{1}{\prod_i k_1!}\frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty d\lambda\ \lambda^{\alpha+k-1}e^{-\lambda\,/\,[\beta/(1+n\beta)]}\\
&= \frac{1}{\prod_i k_1!}\frac{1}{\beta^\alpha\Gamma(\alpha)}\left(\frac{\beta}{1+n\beta}\right)^\alpha\Gamma(\alpha+k)\int_0^\infty d\lambda\frac{1}{\left(\frac{\beta}{1+n\beta}\right)^\alpha\Gamma(\alpha+k)}\lambda^{\alpha+k-1}e^{-\lambda\,/\,[\beta/(1+n\beta)]}
\end{aligned}
$$

Where in the pentultimate step, we pulled out the terms that didn't depend on λ, and then we included the extra factors to write the integral of a Gamma distribution with updated parameters. Kolmogrov's 2nd axiom, the integral of a pdf goes to 1, so the evidence becomes:

$$
P(k) = \frac{1}{\prod_i k_1!}\frac{1}{\beta^\alpha\Gamma(\alpha)}\left(\frac{\beta}{1+n\beta}\right)^\alpha\Gamma(\alpha+k)
$$

Putting the pieces together for the posterior

$$
\begin{aligned}
P(\lambda\,|\,k) &= \frac{P(k\,|\,\lambda)\ P(\lambda;\alpha,\beta)}{P(k)}\\
&= \frac{1}{\frac{1}{\prod_i k_1!}\frac{1}{\beta^\alpha\Gamma(\alpha)}\left(\frac{\beta}{1+n\beta}\right)^\alpha\Gamma(\alpha+k)}\frac{1}{\prod_i k_1!}\frac{1}{\beta^\alpha\Gamma(\alpha)}\lambda^{\alpha+k-1}e^{-\lambda/[\ \beta\,/\,(1+n\beta)\ ]}\\
&= \frac{1}{\left(\frac{\beta}{1+n\beta}\right)^\alpha\Gamma(\alpha+k)}\frac{1}{\lambda^{\alpha+k-1}e^{-\lambda/[\beta/(1+n\beta)]}}
\end{aligned}
$$

2c) Is the posterior in the same functional family as the prior??

Yes! The posterior also follows a Gamma distribution with updated parameters:

$$\alpha^* = k + \alpha = \sum_i k_k + \alpha$$
$$\beta^* = \frac{\beta}{1 + n\beta}$$

Note: If you chose to work through the maths using the exponential or χ^2 prior for $p(\lambda)$, the result would not still be in the exp / χ^2 family, but you would find that the posterior is in the Gamma functional family, which then would then inform that the conjugate prior we were searching for in (a) was the Gamma distribution 😁

2d) What is the Bayes MAP estimator for λ?
(**Helpful:** mode of the Gamma distribuion is $(\alpha - 1)\beta$)

The MAP estimator is the ***mode*** of the posterior, and we can read it off from the Gamma posterior we found in parts (b,c).
(Recall, we stated in class that the mean of the Gamma distribution $P(x; \alpha, \beta)$ is $\mathbb{E}[x] = \alpha\beta$.

$$\hat{p}_{\text{MAP}} = \alpha^*\beta^* = (\alpha + k - 1)\frac{\beta}{1 + n\beta}$$

2e) Show that this MAP estimator can be written as a weighted sum of (1) the mode from the prior and (2) the MLE of λ from the Poisson, $\hat{p}_{\text{MLE}} = k/n$. Do the wieghts that you found make sense?

$$\hat{p}_{\text{MAP}} = (\alpha^* - 1)\beta^* = \frac{1}{1 + n\beta} \cdot (\alpha - 1)\beta \; + \; \frac{\beta n}{1 + n\beta} \cdot \frac{k}{n}$$

- For n = 0 (no data) the MAP estimate of p is just the prior mode $(\alpha - 1)\beta$.
- As $n \to \infty$, the weight of the prior term goes to 0 and $\frac{\beta n}{1 + \beta n} \to 1$ and we recover the MLE estimate from the likelihood.

This is similar to what we saw in class with the Beta prior / binomial likelihoood problem!