UNIVERSITY OF LONDON

Goldsmiths
UNIVERSITY OF LONDON

Probability Density of Parameters and Bayesian Estimation [1]

Math Notes | Larry Cui

April 21, 2022

Even can we find the right pdf model for a whole population of the data, we may never find the true parameters for such models. We construct different kinds of functions to evaluate the parameters, which we call point estimators. We feed sample data to the function and get the result, i.e., estimate(s), for the true parameter(s). If we have different set of samples, apparently we would have different estimates.

The spread out of the estimates is itself a probability density. If the function/estimator is simply the sum or mean of the sample, we know from the CLT that estimates will follow a normal distribution pattern, with a mean of μ and variance of $\sigma/\sqrt{n}$.

When new data sample flows in, it presents an opportunity to check and update our estimates to catch up with the development of whole data conveyed by such new information. Do we need to combine "old" and "new" data and use bothersome maximum likelihood or moments methods to re-calculate the estimates? Yes, you can do that. But we have a better solution with help from Bayesian estimation, and substitution of normal with new pdf model for parameter estimates.

1 Review of Gamma Function

Gamma function is denoted as,

$$\Gamma(r) = \int_0^\infty y^{r-1} e^{-y}\, dy$$

Gamma function has some interesting features. $\Gamma(1) = 1$, $\Gamma(r) = (r-1)\Gamma(r-1)$, and if r is an integer, then $\Gamma(r) = (r-1)!$

Gamma distribution function is different from Gamma function, though the names are a bit

[1] I referred to three blogs from Ms. Aerin Kim:
https://towardsdatascience.com/bayesian-inference-intuition-and-example-148fd8fb95d6
https://towardsdatascience.com/conjugate-prior-explained-75957dc80bfb
https://towardsdatascience.com/beta-distribution-intuition-examples-and-derivation-cf00f4db57af,
and one from Mr. Jonny Brooks Bartlett:
https://towardsdatascience.com/probability-concepts-explained-bayesian-inference-for-parameter-estimation-90e8930e5348

confusing. A random variable Y is said to have the gamma pdf with parameters r and λ if

$$f_Y(y) = \underbrace{\frac{\lambda^r}{\Gamma(r)}}_{\text{constant}} y^{r-1}e^{-\lambda y}, \quad y \geqslant 0$$

Also, as explained in previous notes, $E(Y) = r/\lambda$ and $\operatorname{Var}(Y) = r/\lambda^2$. We need to differentiate the gamma pdf to get the mode: $(r-1)/\lambda$.

2 Brief Introduction to Beta pdf

2.1 Beta Function

The beta function, also called the Euler integral of the first kind, is a special function that is closely related to the gamma function and to binomial coefficients. It is defined by the integral

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$$

A key property of the beta function is its close relationship to the gamma function:

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

Proof To derive this relation, write the product of two factorials as

$$\begin{aligned}\Gamma(x)\Gamma(y) &= \int_{u=0}^{\infty} e^{-u}u^{x-1}\,du \cdot \int_{v=0}^{\infty} e^{-v}v^{y-1}\,dv \\ &= \int_{u=0}^{\infty}\int_{v=0}^{\infty} e^{-u-v}u^{x-1}v^{y-1}\,dv\,du\end{aligned}$$

Let $u = zt$ and $v = z - zt$ to produce

$$\begin{aligned}\Gamma(x)\Gamma(y) &= \int_{z=0}^{\infty}\int_{t=0}^{1} e^{-z}(zt)^{x-1}(z(1-t))^{y-1}z\,dt\,dz && \triangleright\ du{=}zdt \\ &= \int_{z=0}^{\infty} e^{-z}z^{x+y-1}\,dz \cdot \int_{t=0}^{1} t^{x-1}(1-t)^{y-1}\,dt \\ &= \Gamma(x+y)\cdot B(x, y)\end{aligned}$$

Dividing both sides by $\Gamma(x+y)$ gives the desired result.

Because of its relationship to gamma function, beta function can also be denoted as binomial coefficients. When x, y are positive integers,

$$B(x, y) = \frac{(x-1)!(y-1)!}{(x+y-1)!} = \frac{x+y}{xy}\cdot\binom{x+y}{x}^{-1}$$

We can develop some interesting features based on binomial coefficients:

$$B(x+1, y) = \frac{x}{x+y}\cdot\frac{(x-1)!(y-1)!}{(x+y-1)!} = \frac{x}{x+y}\cdot B(x, y)$$

and

$$B(x, y+1) = \frac{y}{x+y} \cdot \frac{(x-1)!(y-1)!}{(x+y-1)!} = \frac{y}{x+y} \cdot B(x, y)$$

and

$$B(x, y) = B(x+1, y) + B(x, y+1)$$

Sometimes we also encounter incomplete beta function, a generalization of the beta function, which is defined as (we use α, β instead of x, y here)

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

When $x = 1$, the incomplete beta function coincides with the complete beta function. Incomplete beta function and its related variants have many applications, but we won't dig further since it's not the topic of this note.

2.2 Beta Distribution

Caution! Beta distribution is related to beta function, but they are different.

According to Wikipedia, in probability theory and statistics, the beta distribution is a family of continuous probability distributions defined on the interval $[0, 1]$ parameterized by two positive shape parameters, denoted by α and β, that appear as exponents of the random variable and control the shape of the distribution. The generalization to multiple variables is called a Dirichlet distribution.

> **Beta distribution** widely used as a probability density function (pdf) for a target parameter θ, when $\theta \in [0, 1]$
>
> $$f_\Theta(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1} = \frac{1}{B(\alpha, \beta)} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

By convention, people use "Beta" to denote beta distribution. For example, if it's about the parameter X, denoted as $X \sim \mathbf{Beta}(\alpha, \beta)$.

Look at the second part of the distribution $\theta^{\alpha-1}(1-\theta)^{\beta-1}$, does it ring a bell to you? Yes, it's the inner part of the beta function $B(\alpha, \beta)$. So the $1/B(\alpha, \beta)$ is here to make sure that when integrated from 0 to 1, the total probability of $f_\Theta(\theta)$ sums to 1.

Corresponding to pdf, we must have a cdf for beta distribution,

$$F_\Theta(\theta) = \int_0^\theta \frac{1}{B(\alpha, \beta)} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta = \underbrace{\frac{1}{B(\alpha, \beta)}}_{constant} \cdot \underbrace{\int_0^\theta \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta}_{integral}$$

The integral part of the right-hand side is in fact an incomplete beta function, so cdf can also be written as

$$F_\Theta(\theta) = \frac{B(\theta; \alpha, \beta)}{B(\alpha, \beta)} = I_\theta(\alpha, \beta)$$

here $I_\theta(\alpha, \beta)$ is called regularized incomplete beta function.

Mean of beta distribution:

$$E(\theta) = \frac{\alpha}{\alpha + \beta}$$

Proof From the definition, we have

$$\begin{aligned}
E(\theta) &= \int_0^1 \frac{\theta}{B(\alpha, \beta)} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta \\
&= \frac{1}{B(\alpha, \beta)} \int_0^1 \theta^{\alpha}(1-\theta)^{\beta-1}\, d\theta && \triangleright \text{ integral part is beta w/ } \alpha{+}1 \\
&= \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} \\
&= \frac{\alpha}{\alpha+\beta} \cdot \frac{B(\alpha, \beta)}{B(\alpha, \beta)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}$$

Variance of beta distribution:

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Proof We start by finding $E(\theta^2)$. We know that

$$E(\theta^2) = \frac{1}{B(\alpha, \beta)} \int_0^1 \theta^{\alpha+1}(1-\theta)^{\beta-1}\, d\theta = \frac{B(\alpha+2, \beta)}{B(\alpha, \beta)}$$

Applying binomial coefficients,

$$\frac{B(\alpha+2, \beta)}{B(\alpha, \beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} \cdot \frac{B(\alpha, \beta)}{B(\alpha, \beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}$$

then by equation $\sigma^2 = E(\theta^2) - \mu^2$,

$$\begin{aligned}
\sigma^2 &= \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
&= \frac{\alpha(\alpha+1)(\alpha+\beta) - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}$$

Mode of beta distribution: the mode is the value of θ at which $f_\Theta(\theta)$ achieves its maximum in scope of $[0, 1]$.

$$\theta = \frac{\alpha - 1}{\alpha + \beta - 2}$$

Proof We use first derivative to find it:

$$
\begin{aligned}
f_\Theta(\theta) &= \frac{1}{B(\alpha,\beta)} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1} \\
f_\Theta^{(1)}(\theta) &= \frac{d}{dx}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&= (\alpha-1)\theta^{\alpha-2}(1-\theta)^{\beta-1} - \theta^{\alpha-1}(\beta-1)(1-\theta)^{\beta-2} \\
&= \theta^{\alpha-2}(1-\theta)^{\beta-2}[(\alpha-1)(1-\theta)-(\beta-1)\theta]
\end{aligned}
$$

If $f_\Theta^{(1)}$ will be at 0, it must be the term in the brackets at 0, so we have,

$$
\begin{aligned}
(\alpha-1)(1-\theta)-(\beta-1)\theta = 0 \\
\theta(\alpha-1+\beta-1) = \alpha-1 \\
\theta_{\text{mode}} = \frac{\alpha-1}{\alpha+\beta-2}
\end{aligned}
$$

An Illustration of mean and mode of a beta distribution shows below. Dr. Bognar at the University of Iowa built the calculator for Beta distribution, which is presenting the beta distribution in an aesthetically pleasant way.

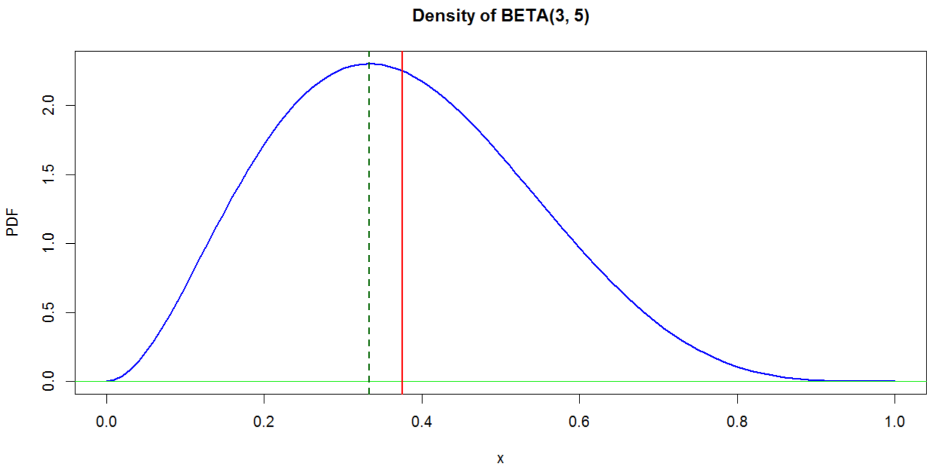


Figure 1: Mean (solid red) and mode (dotted green) of a beta distribution

Comment The other way around: using μ and σ^2 to find α, β [2].
First of all, notice that:

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

This means the variance can be expressed in terms of the mean as

$$\sigma^2 = \frac{\mu(1-\mu)}{\alpha+\beta+1}$$

[2] David Robinson contributed this solution on StackExchange.

so we have $\alpha+\beta$ as

$$\alpha+\beta=\frac{\mu(1-\mu)}{\sigma^2}-1$$

As a result, forms to calculate each parameter

$$\alpha=\mu(\alpha+\beta) \qquad \text{and} \qquad \beta=(1-\mu)(\alpha+\beta)$$

3 Bayes Theorem and Inference

Statistical inference is the process of deducing properties about a population or probability distribution from data. **Bayesian inference** is just the process of deducing properties about a population or probability distribution from data *using Bayes' theorem.*

Much talk about Bayes' theorem has been made, so I posted a picture here for refreshment.

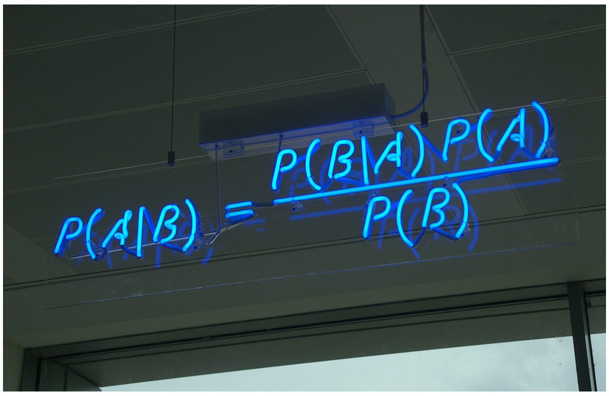


Figure 2: A blue neon sign showing the simple statement of Bayes' theorem

Example 1 We used to apply Bayes to point estimates. Suppose a call center log shows that the incoming call comes in at a rate of 10 calls per unit time (5 min. interval) for 75% of the time, and 8 for the rest 25% of the time.

A sample was taken recently to show the rate is 7 however. How would this discovery change our estimates? We know Poisson is best suitable to model the phone call frequency,

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

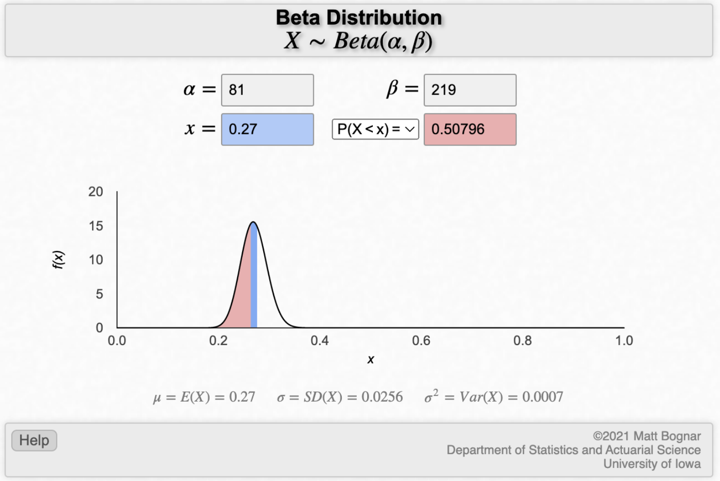
Under assumption of 10 calls, the probability of getting 7 per unit is: $e^{-10}\dfrac{10^7}{7!}$, under 8 is $e^{-8}\dfrac{8^7}{7!}$. We can use Bayes theorem to update our previous estimates:

$$P_{(\lambda=10)}=\frac{75\%\cdot e^{-10}\frac{10^7}{7!}}{75\%\cdot e^{-10}\frac{10^7}{7!}+25\%\cdot e^{-8}\frac{8^7}{7!}}=0.659$$

$$P_{(\lambda=8)}=1-P_{(\lambda=10)}=0.341$$

Comment Up to now, when we talk about probability parameters, we are talking about "values" or "numbers". What if we are not so sure about the parameters, and what we come up with, instead of just a few numbers at our best guess, is *a probability distribution of a parameter of a probability distribution*?

For example, the baseball batting result can be represented with a binomial distribution. The hitting rate p is the only parameter of this pdf, and according to historical record, we estimate $p = 0.27$ with a range from 0.21 to 0.35. We use beta distribution to model the pdf of this parameter p.



Beta Distribution

$X \sim Beta(\alpha, \beta)$

$\alpha =$ 81 $\beta =$ 219

$x =$ 0.27 P(X < x) = 0.50796

$\mu = E(X) = 0.27$ $\sigma = SD(X) = 0.0256$ $\sigma^2 = Var(X) = 0.0007$

Help

©2021 Matt Bognar
Department of Statistics and Actuarial Science
University of Iowa

Figure 3: Parameter is no longer a number but a pdf curve

But can Bayes apply to a parameter pdf?

The answer is "Yes!" First of all, let's see the equation,

$$\overset{\nearrow \text{posterior}}{P(\theta|X)} = \frac{\overset{\nearrow \text{sampling}}{P(X|\theta)} \cdot \overset{\nearrow \text{prior}}{P(\theta)}}{\int_{\underset{\underset{\text{for every possible } \theta}{\downarrow}}{\theta}} P(X|\theta) \cdot P(\theta)\, d\theta} \rightarrow \text{normalizing constant}$$

Figure 4: Bayes Equation - one of the most famous equations in the world of Statistics

Please be noted that we are talking about the probability density function of parameter, here the θ, not the variable from samples or population itself. Bear this in mind.

$P(\theta)$: this is the pdf before we see new evidence (data). We refer to it as the "**prior distribution**". There are some rules to pick the right model for this pdf, as we will discuss later. This is a pdf of data pdf model parameters. Since it's also a pdf, it must come with its own "parameters", DO NOT mix them with the parameters of the data variable pdf! It turns out that the "parameters" of the pdf of parameters are not very important. Sometimes it comes from historical data, sometimes it comes from empirical estimate, and sometimes we can even use a uniform pdf instead. Why? Because prior distribution will evolve and catch up with newly fed data anyway!

$P(X \mid \theta)$: this is the sampling probability. Interpreting in layman words, it is the probability of getting certain x if we pick a value for θ. To be clearer, this probability is calculated based on pdf of variable, NOT the pdf of θ.

$P(\theta \mid X)$: this is the pdf after we apply the Bayes theorem. We refer to it as the "**posterior distribution**". Same as prior distribution, it is still a pdf of parameters.

The denominator part (integral) is called the marginal pdf of X. For a specific x, we iterate every possible θ and pair it with x to find the probability for each pair. Integrating together, we get a constant called "normalizing constant", and this constant is to make sure the total amount is equal to 1, the condition for $P(\theta \mid X)$ still being a pdf.

The work involved in calculating the integral part is usually tedious and messy, and sometimes people are only care about the proportional relationship between $P(\theta \mid X)$ and nominator. So sometimes we can see people just omit the denominator and write the equation as,

$$P(\theta \mid X) \propto P(X \mid \theta) \cdot P(\theta)$$

$\propto$ means "proportional to".

Wait..., we've seen Bayes works from point estimate to point estimate, but how come Bayes also works fine with pdf?

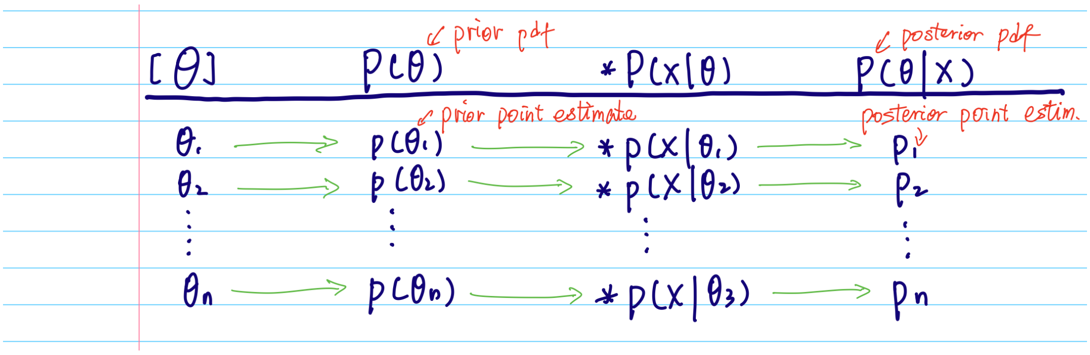


Figure 5: Bayes theorem works fine with pdf

Figure 5 shows that in pdf scenario Bayes still works from point to point. But as we iterate all points/numbers in a pdf domain, what we get as a result range is exactly a group of new points corresponding to the original feed, thus a new form of pdf.

4 Conjugate Prior

What or which prior distribution shall we use to model the parameter pdf? Is there any way to deduce some certain prior distributions from parameters? I haven't found answers to this question, but people do have preference on choosing which distribution as prior.

Let's look at figure 4 again. $P(X \mid \theta)$ is, when we hold X fixed and iterate θ all around the scope, a curve of probability! When we differentiate it to get the $\hat{\theta}$, this is called the **maximum likelihood estimator**. The curve itself, is called **likelihood function**. If we have a sample of size 1, the likelihood function is the variable pdf. If we have a sample of multiple elements, the likelihood function is the variable pdf to the power of multiple number.

For some likelihood functions, if you choose a certain prior, the posterior ends up being in the same kind of distribution as the prior. Such a prior then is called a **Conjugate Prior**. We have the following list of pdf that are conjugate prior to different likelihood/variable pdf(s).

likelihood/variable	prior	posterior
Bernoulli	Beta	Beta
Binomial	Beta	Beta
Negative Binomial	Beta	Beta
Geometric	Beta	Beta
Poisson	Gamma	Gamma
Exponential	Gamma	Gamma
Normal	Normal	Normal

Table 1: Conjugate prior list

A sample is worth a thousand words.

Example 2 ▶ youtube log shows that for each of your post approximately 3% of your viewers will like it, but no more than 6%.

We know this rate, say, θ, would be better represented by a pdf. We also want to update this θ to cope with the latest data we can observe from time to time. As the variable pdf is obviously a binomial distribution, we decide to use a beta pdf to model the value of θ.

We will discuss how to pick α, β to construct this beta pdf for θ, but for now let's just use these Greek letters,

$$f_\Theta(\theta) = \frac{1}{Beta(\alpha, \beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Now we have a sample of size n on your lastest post. It shows that among n viewers k likes your new post. Apparently, what ever the θ is, the probability that k out of n likes your post is

$$P_X(k \mid \theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$$

How do we update the beta pdf? Well, we use the famous Bayes Equation as illustrated on figure 4! Let $g_\Theta(\theta)$ denote posterior distribution,

$$\begin{aligned}
g_\Theta(\theta) &= \frac{P_X(k \mid \theta) f_\Theta(\theta)}{\displaystyle\int_0^1 P_X(k \mid \theta) f_\Theta(\theta)\, d\theta} \\
&= \frac{\binom{n}{k} Beta(\alpha,\beta)^{-1} \cdot \theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}}{\displaystyle\int_0^1 \binom{n}{k} Beta(\alpha,\beta)^{-1} \cdot \theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}\, d\theta} \\
&= \frac{\theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}}{\displaystyle\int_0^1 \theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}\, d\theta} && \triangleright\ \tbinom{n}{k} Beta(\alpha,\beta)^{-1} \text{ constant cancelled out} \\
&= \frac{1}{Beta(k+\alpha, n-k+\beta)} \cdot \theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}
\end{aligned}$$

What do you find? The result is simply add k to α, and add $n-k$ to β. Now you see the power of conjugate prior. Instead of tedious computation, we know the rule and can write the posterior pdf directly!

Comment We said we will discuss how to pick values for α, β for the prior, now let's do it. In practice, we usually equal the expected value "mean" of beta pdf $\alpha/(\alpha+\beta)$ to our safest bet on the parameter. so we have $\alpha/(\alpha+\beta) = 0.03$.

Furthermore, if we have both $\alpha > 1$ and $\beta > 1$, we know the beta pdf is showing a bell curve feature. As a rule of thumb, 3σ radius around mean is enough to encompass the whole range of our guess. so we have $3\sigma = 6\% - 3\% = 3\%$, or $\sigma = 1\%$.

We've discussed how to find α, β from μ, σ. Rounding up, we have $\alpha = 9$, $\beta = 281$.

Example 3 Let's try another example on gamma pdf.

The random variable X is from a Poisson distribution. Usually we use letter λ as the unit rate for such distribution, but here we use θ instead. If we want to model the probability of k in a total of n units, the rate would be $n\theta$:

$$p(k) = \frac{e^{-n\theta}(n\theta)^k}{k!}$$

In reality, we must have some prior experience on the values of $theta$, and can use that estimate to pick the right parameter values for λ and r to construct a gamma pdf. But here for illustration purpose only, let's omit the process of number picking and directly assume the prior gamma pdf for parameter θ be,

$$f_\Theta(\theta) = \frac{\lambda^r}{\Gamma(r)}\theta^{r-1}e^{-\lambda\theta}$$

then we have

$$\begin{aligned} p(k)f_\Theta(\theta) &= \frac{e^{-n\theta}(n\theta)^k}{k!}\frac{\lambda^r}{\Gamma(r)}\theta^{r-1}e^{-\lambda\theta} \\ &= \frac{n^k}{k!}\frac{\lambda^r}{\Gamma(r)}\theta^{k+r-1}e^{-(\lambda+n)\theta} \end{aligned}$$

The posterior distribution is

$$\begin{aligned} g_\Theta(\theta) &= \frac{\frac{n^k}{k!}\frac{\lambda^r}{\Gamma(r)}\theta^{k+r-1}e^{-(\lambda+n)\theta}}{\int_0^\infty \frac{n^k}{k!}\frac{\lambda^r}{\Gamma(r)}\theta^{k+r-1}e^{-(\lambda+n)\theta}\,d\theta} \\ &= \frac{\frac{n^k}{k!}\frac{\lambda^r}{\Gamma(r)}}{\frac{n^k}{k!}\frac{\lambda^r}{\Gamma(r)}\int_0^\infty \theta^{k+r-1}e^{-(\lambda+n)\theta}\,d\theta}\cdot\theta^{k+r-1}e^{-(\lambda+n)\theta} \\ &= \frac{1}{\int_0^\infty \theta^{k+r-1}e^{-(\lambda+n)\theta}\,d\theta}\cdot\theta^{k+r-1}e^{-(\lambda+n)\theta} \end{aligned}$$

Let $t = (\lambda+n)\theta$, the above denominator is transformed to

$$\begin{aligned}\int_0^\infty \theta^{k+r-1} e^{-(\lambda+n)\theta}\, d\theta &= \int_0^\infty \left(\frac{t}{\lambda+n}\right)^{k+r-1} e^{-t}\frac{1}{\lambda+n}\, dt \\ &= \frac{1}{(\lambda+n)^{k+r}} \int_0^\infty t^{k+r-1} e^{-t}\, dt && \triangleright \text{ Integral part is a gamma func.} \\ &= \frac{\Gamma(k+r)}{(\lambda+n)^{k+r}}\end{aligned}$$

Posterior distribution can therefore be written as:

$$g_\Theta(\theta) = \frac{(\lambda+n)^{k+r}}{\Gamma(r)} \cdot \theta^{k+r-1} e^{-(\lambda+n)\theta}$$

a neat transformation from conjugate prior $f_\Theta(\theta)$.

5 MAP and Bayesian Estimation

Once we have a posterior distribution, how can we find our estimate for the parameter? One approach, similar to using the likelihood function to find a maximum likelihood estimator, is to differentiate the posterior, in which case the value for $dg_\Theta \,/\, d\theta = 0$. θ value at this point is called “mode”. The way to find mode of posterior has a term: maximum a posteriori probability (MAP) estimate.

Another way is based on minimizing risk associated with estimated $\hat{\theta}$, where the risk is the expected value of a loss function. Two of the most frequently used loss functions are:

$$L(\hat{\theta}, \theta) = \begin{cases} \left|\hat{\theta} - \theta\right| \\ (\hat{\theta} - \theta)^2 \end{cases}$$

then

$$\textsc{Risk} = \int_\theta L(\hat{\theta}, \theta) g_\Theta(\theta)\, d\theta$$

$\hat{\theta}$ is picked from the domain of posterior once we have the minimum risk. It can be proved that for absolute loss function $\hat{\theta}$ = median, and for square $\hat{\theta}$ = mean. [LM12, p. 339]

Reference

[LM12] R.J. Larsen and M.L. Marx. *An Introduction to Mathematical Statistics and Its Applications*. Prentice Hall, 2012. ISBN: 9780321693945. URL: https://books.google.com.hk/books?id=tZdbRAAACAAJ.