Simple Bayesian Computations

Marshall Bradley

Bayesian probability theory is the logically consistent calculus for inference in the presence of uncertainty. The Bayesian approach requires us to express our knowledge of unknown parameters in a model in the form of a prior probability distribution. We must also express the relationship between the measurements we make and the parameters in the form of a likelihood function. This likelihood function is a probability distribution with respect to the data but not the parameters. In simple cases involving independent measurements it is the product of the sampling function. Bayes theorem tells us that the posterior probability distribution is proportional to the product of the prior and the likelihood. The constant of proportionality is called the evidence and it is a multidimensional integral. Here we present examples in which the number of dimensions is one or two and all necessary computations can either be done analytically or by brute force numerical integration.

A Bayesian example from actuarial science

References

(1) Klugman, Stuart A. (1992), *Bayesian Statistics in Actuarial Sciences*, Kluwer Academic Publishers.
(2) Skilling, John (2010), "Foundations and algorithms", published in *Bayesian Methods in Cosmology*, Cambridge University Press edited by Michael P. Hobson et al.

The complete calculus of inference can be summarized by the following three equations (Skilling, 2010). We begin with a prior probability distribution $\pi(\theta)$ that characterizes our logically initial state of knowledge about some unknown parameter θ. We observe some data D that are dependent upon the parameter θ. The relationship between data and parameter is characterized by a likelihood function $L(D \mid \theta)$. Bayes theorem tell us that the posterior distribution of the parameter θ conditioned upon the knowledge of the data is

$$P(\theta \mid D) = \frac{1}{E} \pi(\theta) L(D \mid \theta)$$

where the evidence E is the normalizing factor that ensures that the posterior is a normalized probability distribution:

$$E = \int_{\theta_{\min}}^{\theta_{\max}} \pi(\theta) L(D \mid \theta) d\theta$$

Additionally it must be the case that the prior is properly normalized so that

$$\int_{\theta_{\min}}^{\theta_{\max}} \pi(\theta) d\theta = 1$$

The evidence represents how well our original assessment of θ manages to predict the data. The evidence is sometimes referred to as the prior predictive or the marginal likelihood. However, evidence is a better descriptor since it plays such a central role in Bayesian data analysis.

We consider a specific numerical example from the book by Stuart Klugman entitled *Bayesian Statistics in Actuarial Sciences*. Suppose we consider the number of insurance claims from one randomly selected driver from the collection of all insured drivers. Assume that the number of claims from one driver follows a Poisson distribution with parameter θ and that the number of claims in different years are independent. If the observed data D are $x_1\, x_2 \,..\, x_n$ then the likelihood of the data is

$$L(D \mid \theta) = \prod_{i=1}^{n} \frac{\exp(-\theta)\, \theta^{x_i}}{x_i!} = \frac{1}{\prod_{i=1}^{n} x_i} \exp(-n\theta)\, \theta^{\sum_{i=1}^{n} x_i}$$

We will assume that the prior distribution on the unknown parameter θ is gamma distributed with parameters a and b:

$$\pi(\theta) = \frac{b^{-a}\, e^{-\frac{\theta}{b}}\, \theta^{-1+a}}{\text{Gamma}[a]}$$

We will make the specific choice that $a = 0.1$ and $b = 3$. This choice of a and b puts lots of probability mass near $\theta = 0$.

The prior probability distribution and its plot are:

```
In[◦]:= Clear[θ, a, b];
prior[θ_, a_, b_] := (b^-a ⅇ^(-θ/b) θ^(-1+a)) / Gamma[a]
With[{a = 0.1, b = 3}, Plot[prior[θ, a, b], {θ, 0.1, 3},
  PlotRange → All, Axes → False, Frame → True, FrameLabel → {⋯}]]
```

Out[◦]=

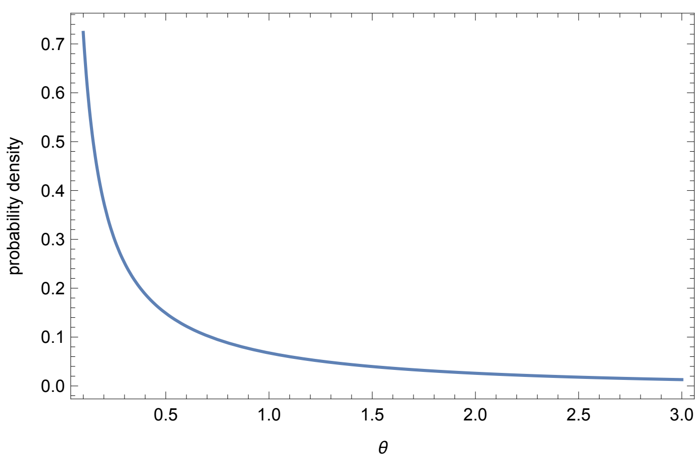

Figure 1. The gamma prior with shape $a = 0.1$ and scale $b = 3$.

The specific data we consider are $x = \{0, 1, 0, 2, 0, 0, 0, 1, 0, 0\}$.

Our likelihood function is defined by:

```
In[◦]:= Clear[θ, n, Σx, Q];
L[θ_, n_, Σx_, Q_] := (1/Q) Exp[-n * θ] θ^Σx
```

In our likelihood function the constant Q is defined via

$$Q = \prod_{i=1}^{n} x_i\,!$$

The quantity Q depends on the data but it does not depend on the unknown parameter θ or the prior parameters a and b. Thus the posterior probability distribution will not depend on Q but the evidence will.

The analytic value of the evidence is :

```
In[∘]:= Integrate[L[θ, n, Σx, Q] × prior[θ, a, b],
          {θ, 0, ∞}, Assumptions → {n ≥ 1, a > 0, b > 1, Σx > 0}]
```

Out[∘]=

$$\frac{b^{\Sigma x}\,(1+b\,n)^{-a-\Sigma x}\,\text{Gamma}[a+\Sigma x]}{Q\,\text{Gamma}[a]}$$

So we define the function:

In[∘]:= $\text{evidence}[a\_, b\_, n\_, \Sigma x\_, Q\_] := \frac{b^{\Sigma x}\,(1+b\,n)^{-a-\Sigma x}\,\text{Gamma}[a+\Sigma x]}{Q\,\text{Gamma}[a]}$

The analytic form of the posterior is:

```
In[∘]:= FullSimplify[L[θ, n, Σx, Q] × prior[θ, a, b] / evidence[a, b, n, Σx, Q]]
```

Out[∘]=

$$\frac{b^{-a-\Sigma x}\,e^{-\left(\left(\frac{1}{b}+n\right)\theta\right)}\,(1+b\,n)^{a+\Sigma x}\,\theta^{-1+a+\Sigma x}}{\text{Gamma}[a+\Sigma x]}$$

So we define the function:

In[∘]:= $\text{posterior}[\theta\_, a\_, b\_, n\_, \Sigma x\_] := \frac{b^{-a-\Sigma x}\,e^{-\left(\left(\frac{1}{b}+n\right)\theta\right)}\,(1+b\,n)^{a+\Sigma x}\,\theta^{-1+a+\Sigma x}}{\text{Gamma}[a+\Sigma x]}$

The posterior must integrate to unity. We check this:

```
In[∘]:= Integrate[posterior[θ, a, b, n, Σx],
          {θ, 0, ∞}, Assumptions → {n ≥ 1, a > 0, b > 1, Σx > 0}]
Out[∘]=
1
```

Now we can plot the posterior probability density function of the parameter θ :

In[∘]:=
```
Module[{a = 0.1, b = 3, x, Σx, n, Q},
 x = {0, 1, 0, 2, 0, 0, 0, 1, 0, 0};
 Σx = Total[x]; n = Length[x];
 Q = Product[Factorial[x[[i]]], {i, 1, n}];
 Plot[posterior[θ, a, b, n, Σx], {θ, 0, 3},
  PlotRange → All, Axes → False, Frame → True, ⋯ → ⋯]
]
```

Out[∘]=

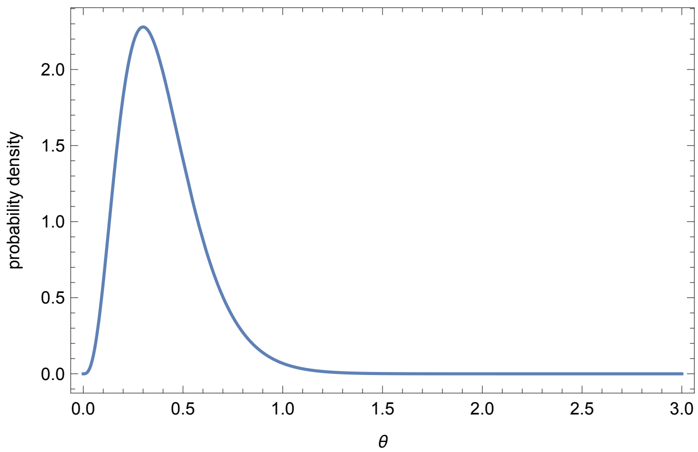

Figure 2. Posterior probability distribution for the Poisson parameter θ.

We compare the analytic evaluation of the evidence with a numerical evaluation of the evidence:

In[∘]:=
```
Module[{a, b, x, Σx, n, Q, analytic, numerical},
 a = 0.1; b = 3;
 x = {0, 1, 0, 2, 0, 0, 0, 1, 0, 0};
 Σx = Total[x]; n = Length[x];
 Q = Product[Factorial[x[[i]]], {i, 1, n}];
 analytic = evidence[a, b, n, Σx, Q];
 numerical = NIntegrate[prior[θ, a, b] × L[θ, n, Σx, Q], {θ, 0, 20}];
 {analytic, numerical}]
```

Out[∘]=

```
{0.0000222764, 0.0000222764}
```

They are identical.

The analytic form posterior predictive distribution for the number of claims in one year is a weighted sum of data likelihoods:

In[ ]:=

```
Integrate[1/Gamma[y + 1] Exp[-θ] θ^y posterior[θ, a, b, n, Σx],
 {θ, 0, ∞}, Assumptions → {n ≥ 1, a > 0, b > 1, Σx > 0, y ≥ 0}]
```

Out[ ]=

$$\frac{\left(\frac{1}{b}+n\right)^{a+\Sigma x}\left(1+\frac{1}{b}+n\right)^{-a-y-\Sigma x}\text{Gamma}[a+y+\Sigma x]}{\text{Gamma}[1+y]\,\text{Gamma}[a+\Sigma x]}$$

So we define:

In[ ]:=

```
posteriorPredictive[y_, a_, b_, n_, Σx_] := ((1/b + n)^(a+Σx) (1 + 1/b + n)^(-a-y-Σx) Gamma[a + y + Σx]) / (Gamma[1 + y] Gamma[a + Σx])
```

And then we plot the posterior predictive distribution :

In[ ]:=

```
Module[{a, b, x, Σx, n, Q, analytic, numerical},
 a = 0.1; b = 3;
 x = {0, 1, 0, 2, 0, 0, 0, 1, 0, 0};
 Σx = Total[x]; n = Length[x];
 DiscretePlot[posteriorPredictive[y, a, b, n, Σx],
  {y, 0, 5}, Axes → False, Frame → True, FrameLabel → {⋯}]]
```

Out[ ]=

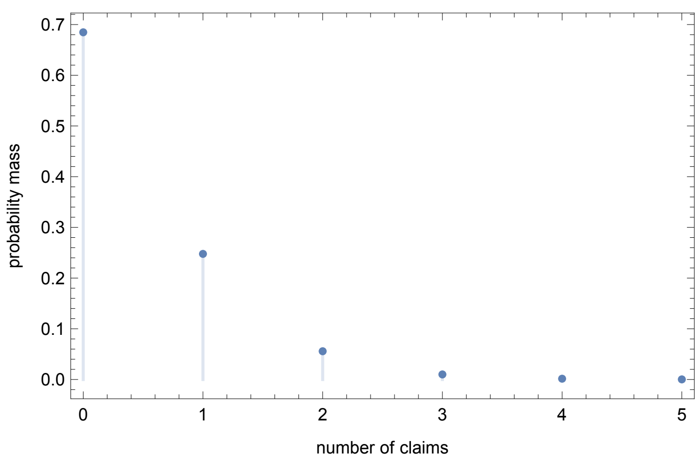

Figure 3. Posterior predictive distribution for the number of claims in one year.

The mean of the posterior predictive distribution is:

In[ ]:=

```
Module[{a, b, x, Σx, n, Q, analytic, numerical},
 a = 0.1; b = 3;
 x = {0, 1, 0, 2, 0, 0, 0, 1, 0, 0};
 Σx = Total[x]; n = Length[x];
 Sum[y * posteriorPredictive[y, a, b, n, Σx], {y, 0, 20, 1}]
]
```

Out[ ]=

```
0.396774
```

Finding the scale parameter of an exponential distribution

Suppose that we want to determine the scale parameter λ in an exponential distribution with probability density function defined by

$$p(x \mid \lambda) = \frac{1}{\lambda} \exp\{-x/\lambda\}, \; x > 0$$

The determination of λ will be based upon observing the data $D: x_1 x_2 .. x_n$.

The prior probability of λ is taken to vary over the range $\lambda_1 < \lambda < \lambda_2$ in accordance with the prior probability distribution $f(\lambda) \propto 1/\lambda$. This is a Jeffreys prior named in honor of Sir Harold Jeffreys, one of the Bayesian giants of the 20th century. Then on prior information H we have apart from a normalization constant $1/\log(\lambda_2/\lambda_1)$

$$P(d\lambda \mid H) = f(\lambda)\, d\lambda = \frac{d\lambda}{\lambda}, \; \lambda_1 < \lambda < \lambda_2.$$

The probability given λ that an observed value of x will lie in the range $(x, x + dx)$ is

$$P(dx \mid \lambda H) = \frac{1}{\lambda} \exp\{-x/\lambda\}\, dx.$$

If the observed data values $D: x_1 x_2 \ldots x_n$ are independent then the joint probability of the data is found by multiplication. Thus

$$P(dx_1 \; dx_2 \ldots dx_n \mid \lambda H) = \frac{1}{\lambda^n} \exp\{-\sum_{i=1}^{n} x_i/\lambda\}\, dx_1\, dx_2 \ldots dx_n.$$

If we define the sufficient statistic $\overline{x}$ via

$$n\,\overline{x} = \sum_{i=1}^{n} x_i$$

then Bayes theorem tells us that

$$P(d\lambda \mid x_1 x_2 \ldots x_n H) \propto f(\lambda) \frac{1}{\lambda^n} \exp\{-n\,\overline{x}/\lambda\}\, d\lambda \propto \frac{1}{\lambda^{n+1}} \exp\{-n\,\overline{x}/\lambda\}\, d\lambda$$

The evidence of the data is proportional to

$$Z = \int_{\lambda_1}^{\lambda_2} \frac{1}{\lambda^{n+1}} \exp\{-n\,\overline{x}/\lambda\}\, d\lambda$$

The posterior probability distribution is

$$p(\lambda \mid n\overline{x}\, H) = \frac{1}{Z\lambda^{n+1}} \exp\{-n\,\overline{x}/\lambda\}, \; \lambda_1 < \lambda < \lambda_2$$

In[ ]:=

Numerical example

The following code will generate synthetic data and compute the posterior with a user input of the true value of λ:

In[◦]:=
```
Clear[posterior, n, λtrue];
posterior[n_, λtrue_] := Quiet@Module[{λ1 = 0.1, λ2 = 30, x, nxbar, Z, pdflocal},
    SeedRandom[1239];
    x = RandomVariate[ExponentialDistribution[1 / λtrue], {n}];
    dataout = Transpose[{x, Table[0.0, Length[x]]}];
    nxbar = Total[x];
    Z = NIntegrate[1/λ^(n+1) Exp[-nxbar / λ], {λ, λ1, λ2}];
    Table[{λ, 1/(Z * λ^(n+1)) Exp[-nxbar / λ]}, {λ, λ1, λ2, (λ2 - λ1)/200}]];
```

If the true value of λ is 7 then the posterior looks like the following:

In[◦]:=
```
λtrue = 7;
ListLinePlot[{posterior[3, λtrue], posterior[7, λtrue],
  posterior[11, λtrue], posterior[21, λtrue]}, PlotRange → All, Axes → False,
 Frame → True, FrameLabel → {"λ: scale", "probability density"},
 PlotStyle → {⋯}, PlotLegends → {⋯}, Epilog → {⋯}]
```

Out[◦]=

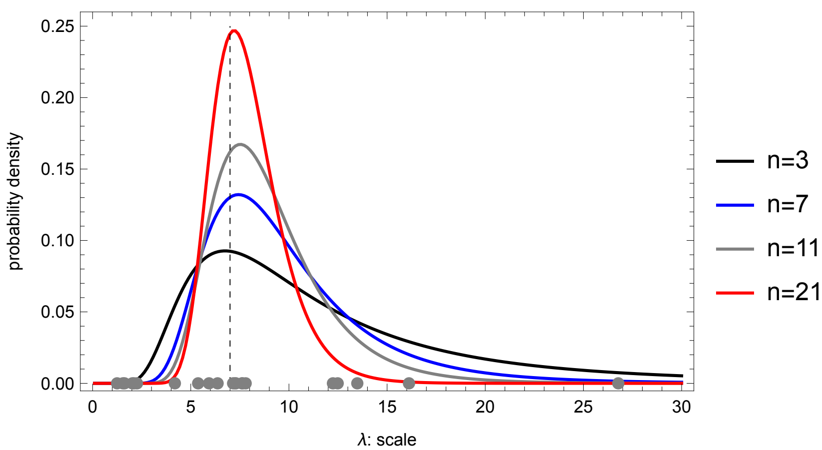

Figure 1. Posterior inference of the scale parameter for an exponential distribution for various sample sizes. The true parameter value is λ=7 as indicated by the vertical dotted line. Each curve is a probability density function. Note that as the number of samples *n* increases, that the posterior becomes more peaked.

Two dimensional Bayesian example

Suppose that we want to determine the scale parameter λ and the shape parameter α of a gamma distribution with probability density function defined by

$$p(x \mid \lambda) = \frac{1}{\Gamma(\alpha)\,\lambda} \left(\frac{x}{\lambda}\right)^{\alpha-1} \exp\{-x/\lambda\}, \; x > 0$$

where $\Gamma(\alpha) = (\alpha - 1)!$. The determination of λ and α will be based upon observing the data $D: x_1\, x_2 \,..\, x_n$.

The prior probability of λ is taken to vary over the range $\lambda_1 < \lambda < \lambda_2$ in accordance with the prior probability distribution $f(\lambda) = 1/\lambda$. Then on prior information H we have (apart from a normalization constant $1/\log(\lambda_2/\lambda_1)$ which is only important in an evidence computation)

$$P(d\lambda \mid H) = f(\lambda)\, d\lambda = \frac{d\lambda}{\lambda},\ \lambda_1 < \lambda < \lambda_2$$

We will assume a flat prior for α over the range $\alpha_1 < \alpha < \alpha_2$. Then on prior information H we have

$$P(d\alpha \mid H) = g(\alpha)\, d\alpha = \frac{d\alpha}{\alpha_2 - \alpha_1},\ \alpha_1 < \alpha < \alpha_2$$

If the observed data values $D:\ x_1\, x_2\, ..\, x_n$ are independent then the joint probability of the data is found by multiplication. Thus

$$P(dx_1\ dx_2 \ldots dx_n \mid \lambda\alpha H) = \frac{1}{\Gamma(\alpha)^n\, \lambda^n} \prod_{i=1}^{n} \left(\frac{x_i}{\lambda}\right)^{\alpha-1} \exp\{-\textstyle\sum_{i=1}^{n} x_i/\lambda\}\, dx_1\, dx_2 \ldots dx_n.$$

This can be written in the simplified form

$$P(dx_1\ dx_2 \ldots dx_n \mid \lambda\alpha H) = \frac{1}{\Gamma(\alpha)^n\, \lambda^{n\alpha}} \left(\textstyle\prod_{i=1}^{n} x_i\right)^{\alpha-1} \exp\{-\textstyle\sum_{i=1}^{n} x_i/\lambda\}\, dx_1\, dx_2 \ldots dx_n.$$

If we define the sufficient statistics $\overline{x}$ and x_Π via

$$n\,\overline{x} = \textstyle\sum_{i=1}^{n} x_i, \quad x_\Pi = \prod_{i=1}^{n} x_i$$

then Bayes theorem tells us that

$$P(d\lambda\ d\alpha \mid x_1 x_2\, ..\, x_n\, H) \propto f(\lambda)\, g(\alpha)\, \frac{1}{\Gamma(\alpha)^n\, \lambda^{n\alpha}}\, (x_\Pi)^{\alpha-1} \exp\{-n\,\overline{x}/\lambda\}\, d\lambda\, d\alpha$$

The posterior probability distribution is

$$p(\lambda\alpha \mid x_\Pi\ n\,\vec{x}\, H) = \frac{1}{Z}\, \frac{1}{\alpha_2 - \alpha_1}\, \frac{1}{\lambda}\, \frac{1}{\Gamma(\alpha)^n\, \lambda^{n\alpha}}\, (x_\Pi)^{\alpha-1} \exp\{-n\,\overline{x}/\lambda\}$$

where Z is the evidence.

In[ ]:=

Numerical example

In the following we present a simple numerical example.

An algorithm for generating synthetic data and computing the two dimensional posterior is as follows :

In[∘]:=
```
n = 31;
λtrue = 11; αtrue = 2.5; λ1 = 0.1; λ2 = 30; α1 = 1.0; α2 = 5.0;
Quiet@Module[{},
   SeedRandom[1238];
   x = RandomVariate[GammaDistribution[αtrue, λtrue], {n}];
   dataout = Transpose[{x, Table[0.0, Length[x]]}];
   nxbar = Total[x];
   xpi = Product[x[[i]], {i, 1, n}]; dα = (α2 - α1)/150; dλ = (λ2 - λ1)/200;
   pdflocal = (dα * dλ)/(α2 - α1) 1/Log[λ2 / λ1]
     Table[1/λ 1/(Gamma[α]^n λ^(n*α)) xpi^(α-1) Exp[-nxbar / λ], {α, α1, α2, dα}, {λ, λ1, λ2, dλ}];
   pdflocal = pdflocal / Total[Flatten[pdflocal]]];
```

The plot of the two dimensional posterior surface is:

In[∘]:=
```
ReliefPlot[pdflocal, LightingAngle → None,
 PlotRange → All, DataRange → {{λ1, λ2}, {α1, α2}},
 FrameTicks → True, AspectRatio → 0.5, FrameLabel → {⋯}]
```

Out[∘]=

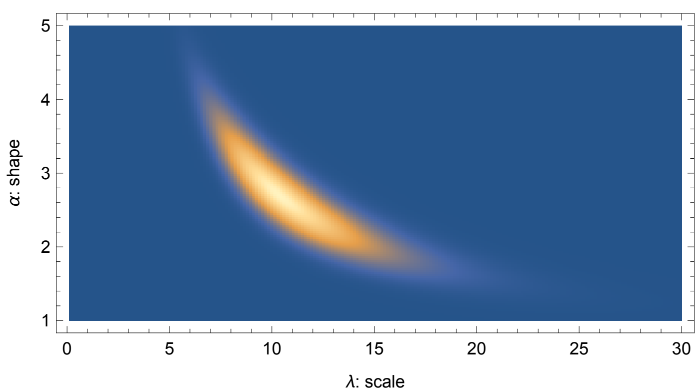

Figure 1. Joint posterior probability distribution.

In the following computations we make sure that the posterior probability function is correctly normalized. Then we sum across and down to obtain the two marginal distributions.

The posterior probability mas should sum to unity:

In[∘]:=
```
Total@Flatten[pdflocal]
```

Out[∘]=
```
1.
```

The marginal posterior distribution of the shape parameter α is shown in the following plot:

In[∘]:=
```
ListLinePlot[Map[Total, pdflocal], DataRange → {α1, α2},
 Axes → False, Frame → True, FrameLabel → {⋯}]
```

Out[∘]=

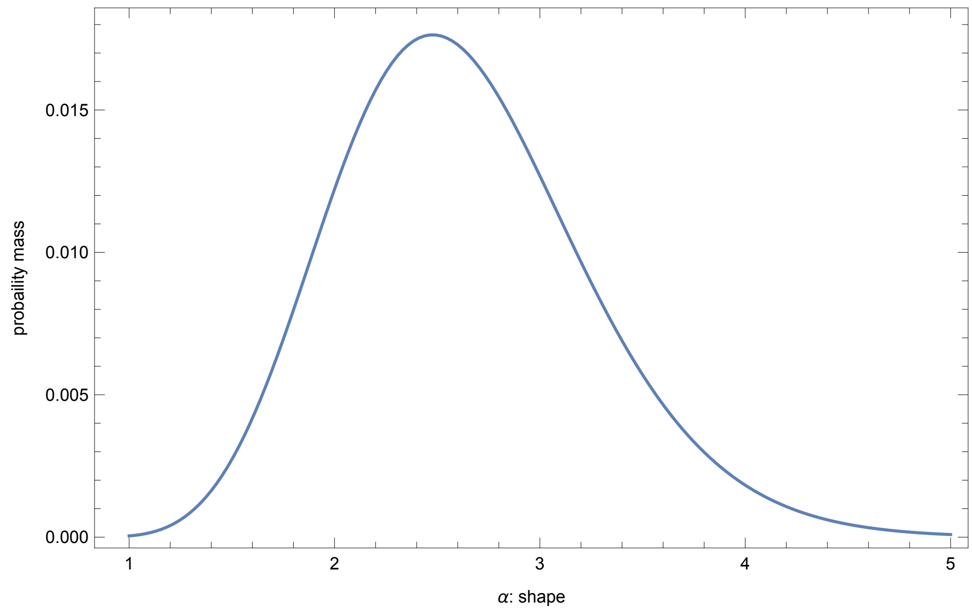

Figure 2. Marginal posterior probability mass function for the shape parameter α. Probability mass and probability density differ here by the factor $1/d\alpha$.

The marginal posterior distribution of the shape scale parameter λ is shown in the following plot:

In[∘]:=
```
ListLinePlot[Map[Total, Transpose[pdflocal]],
 DataRange → {λ1, λ2}, Axes → False, Frame → True, FrameLabel → {⋯}]
```

Out[∘]=

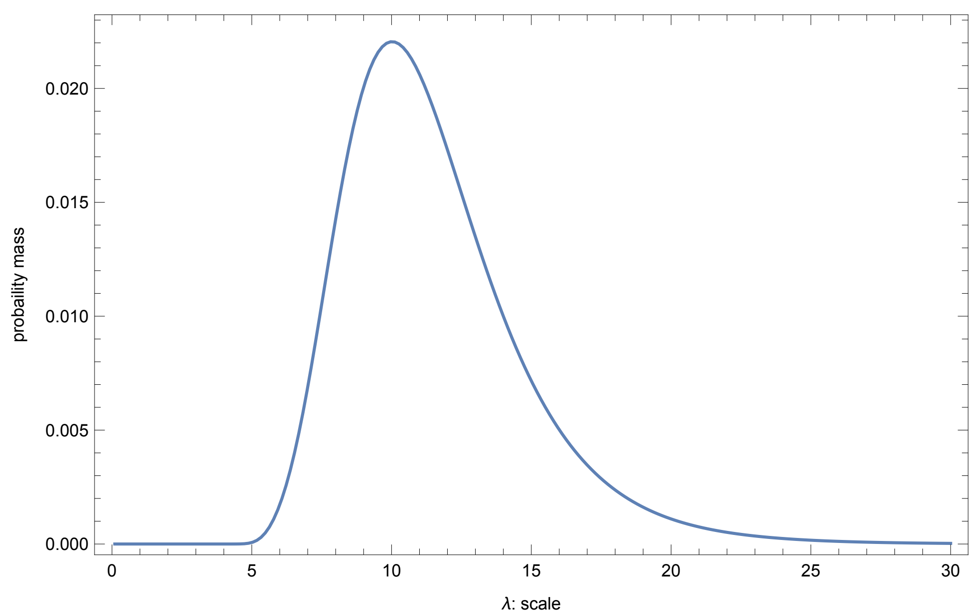

Figure 3. Marginal posterior probability distribution for the scale parameter λ. Probability mass and

probability density differ here by the factor $1/d\lambda$.

Malmquist-Eddington bias

Reference. Hobson, Michael P., Jaffe, Andrew H., Liddle, Andre R., Mukherjee, Pia and Parkinson, David. (2010). *Bayesian Methods in Cosmology*, Cambridge University Press, see figure 12,2.

The idea that we present here is adapted from the work of Stefano Andreon as presented in the book *Bayesian Methods in Cosmology*.

We are interested in estimating the rate μ at which an object emits photons or a detector receives photons. The process of emission is assumed to be governed by a Poisson process with rate μ. We further suppose that we observe $n = 4$ photon arrivals (or counts) in a time period of interest and that based upon prior knowledge the prior follows a power law distribution. The likelihood and the prior are then respectively given by

$$L(n \mid \mu) = \frac{e^{-\mu} \mu^n}{n!}, \quad p(\mu) = C_0 \mu^{-\alpha}$$

where C_0 is a normalization constant that will cancel out in the computation of the posterior probability distribution $p(\mu \mid n)$. The relationship between likelihood, prior and posterior is given by Bayes' theorem.

$$p(\mu \mid n) = \frac{L(n \mid \mu)\, p(\mu)}{\int_0^\infty L(n \mid \mu)\, p(\mu)\, d\mu}$$

We assume a prior probability distribution that is in the form of a power law. This distribution is also referred to as a Pareto distribution if truncated before reaching $\mu = 0$. The normalization constant that we need to compute in order to to find the posterior is called the evidence *E*.

Specifically the evidence is as follows:

In[∘]:=
```
Clear[μ, n, α];
Integrate[Exp[-μ] μ^n/Gamma[1 + n] μ^-α, {μ, 0, ∞}]
```

Out[∘]=

$$\frac{\text{Gamma}[1+n-\alpha]}{\text{Gamma}[1+n]} \quad \text{if } \text{Re}[n-\alpha] > -1$$

The posterior probability density function is:

In[∘]:=
```
Exp[-μ] μ^n/Gamma[1 + n] μ^-α / (Gamma[1 + n - α]/Gamma[1 + n])
```

Out[∘]=

$$\frac{e^{-\mu} \mu^{n-\alpha}}{\text{Gamma}[1+n-\alpha]}$$

The mean value of the posterior distribution is:

In[ ]:=
```
Integrate[μ (ⅇ^-μ μ^(n-α))/Gamma[1 + n - α], {μ, 0, ∞}]
```

Out[ ]=

$$\frac{\text{Gamma}[2+n-\alpha]}{\text{Gamma}[1+n-\alpha]} \quad \text{if} \quad \text{Re}[n-\alpha] > -2$$

With specific parameter choices the mean is:

In[ ]:=
```
Gamma[2 + n - α]/Gamma[1 + n - α] /. {n → 4, α → 2.5}
```

Out[ ]=
```
2.5
```

The prior, likelihood and posterior are shown in the plot below:

In[ ]:=
```
With[{n = 4, α = 2.5},
 Plot[{Callout[μ^-α, "prior", Above], Callout[1/Gamma[n + 1] μ^n Exp[-μ], "likelihood",
    Above], Callout[(ⅇ^-μ μ^(n-α))/Gamma[1 + n - α], "posterior", Above]}, {μ, 0, 10},
  AxesLabel → {⋯}, PlotRange → {{0, 10}, {0, 0.5}}, Epilog → {⋯}]]
```

Out[ ]=

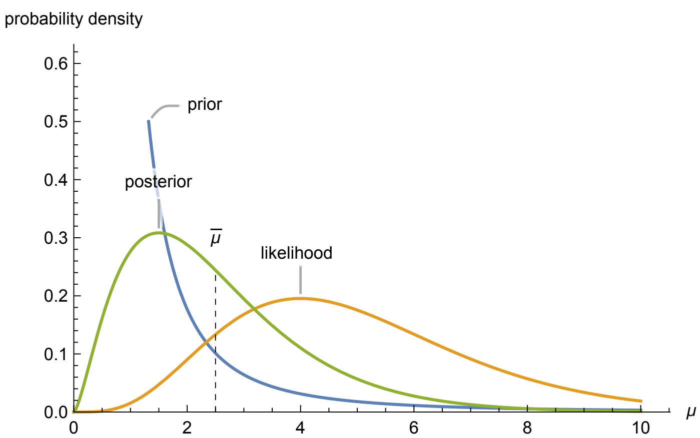

Figure 1. Prior, likelihood and posterior. The posterior mean is indicated by the vertical dotted line.

But we have actually made a subtle but possibly significant mistake. The prior must be a proper probability density function. This means it must be normalizable. If we integrate $1/\mu^{\alpha}$ over $(0, \infty)$. we get an unbounded result. If we integrate over over a finite range of μ we obtain a valid normalization constant

$$C(\mu_1, \mu_2) = \int_{\mu_1}^{\mu_2} \mu^{-\alpha}\, d\mu = \frac{(\mu 1\, \mu 2)^{-\alpha} \left(-\mu 1^{\alpha}\, \mu 2 + \mu 1\, \mu 2^{\alpha}\right)}{-1+\alpha}, \qquad p(\mu) = \frac{1}{C(\mu_1,\mu_2)}\, \mu^{-\alpha} \text{ for } \mu_1 < \mu < \mu_2$$

The posterior is

$$p(\mu \mid n) = \frac{1}{E}\, \frac{1}{C(\mu_1,\mu_2)}\, \mu^{-\alpha}\, \frac{e^{-\mu} \mu^{n}}{n!}, \quad E = \int_{\mu_1}^{\mu_2} \frac{1}{C(\mu_1,\mu_2)}\, \mu^{-\alpha}\, \frac{e^{-\mu} \mu^{n}}{n!}\, d\mu$$

Anything that does not depend on μ cancels and the posterior simplifies to

$$p(\mu \mid n) = \frac{1}{\int_{\mu_1}^{\mu_2} \mu^{-\alpha} e^{-\mu} \mu^n d\mu} \; \mu^{-\alpha} e^{-\mu} \mu^n$$

If we let the range of integration extend over $(0, \infty)$ then the posterior becomes:

In[ ]:=
```
μ^-α Exp[-μ] μ^n / (Integrate[μ^-α Exp[-μ] μ^n, {μ, 0, ∞}])
```

Out[ ]=
$$\frac{e^{-\mu} \mu^{n-\alpha}}{\text{Gamma}[1+n-\alpha]} \quad \text{if } \text{Re}[n-\alpha] > -1$$

This is exactly what we go before, i.e.

$$p(\mu \mid n) = \frac{1}{\text{Gamma}[1+n-\alpha]} \; e^{-\mu} \mu^{n-\alpha}$$

provided that $n - \alpha > -1$. What we do not get is a finite value of the evidence E. The way to avoid this difficulty is to always work over a finite range of the unknown parameter μ.

All of these analytic complexities can be avoided if we choose to work numerically from the onset.

We perform this in the following block of code and compute a plot :

In[ ]:=
```
Module[{α, n, dμ, μ1, μ2, prior, likelihood, posterior},
 α = 2.5; n = 4; dμ = 0.01; μ1 = dμ; μ2 = 10.0;
 prior = Table[μ^-α, {μ, μ1, μ2, dμ}]; prior = prior / Total[prior];
 likelihood = Table[1/Gamma[n + 1] μ^n Exp[-μ], {μ, μ1, μ2, dμ}];
 posterior = Table[e^-μ μ^(n-α) / Gamma[1 + n - α], {μ, μ1, μ2, dμ}];
 posterior = posterior / Total[posterior];
 posterior = posterior / dμ;
ListLinePlot[{100 000 prior, Callout[likelihood, "likelihood", Above],
   Callout[posterior, "posterior", Above]},
  DataRange → {μ1, μ2}, PlotRange → {0, 0.65}, AxesLabel → {⋯}]]
```

Out[ ]=

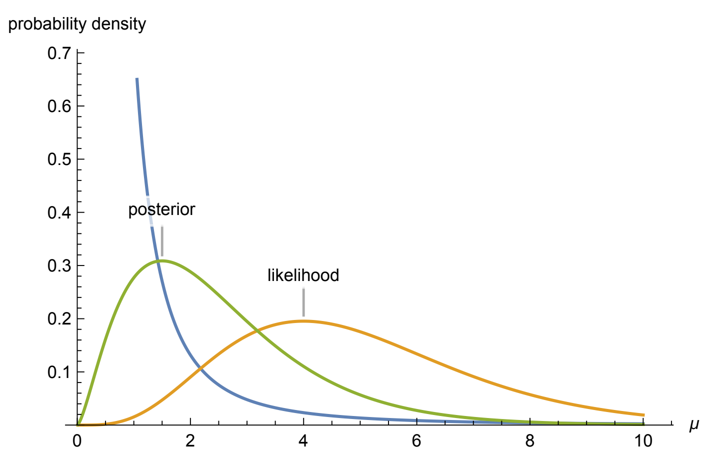

Figure 2. Numerical computation of prior, likelihood and posterior . The prior pdf in the plot has been rescaled for visual purposes.