Gaussian signals exercise – Analytic derivations

April 8, 2025

1 Analytic posterior

We wish to estimate the unknown mean μ and variance σ^2 of a Gaussian signal given n independent samples

$$\mathbf{d}_{\text{obs}} = \{d_{\text{obs},1}, d_{\text{obs},2}, \ldots, d_{\text{obs},n}\}.$$

We assume that, conditioned on μ and σ^2, the data are independent and identically distributed (i.i.d.) Gaussian random variables. In addition, our prior on (μ, σ^2) is taken as a Gaussian-inverse-Gamma,

$$(\mu, \sigma^2) \sim \mathcal{G}\mathit{\Gamma}^{-1}(\eta, \lambda, \alpha, \beta),$$

whose density is

$$\mathcal{G}\mathit{\Gamma}^{-1}(\mu, \sigma^2|\eta, \lambda, \alpha, \beta) = \frac{\sqrt{\lambda}}{\sqrt{2\pi\sigma^2}} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(-\frac{2\beta + \lambda(\mu - \eta)^2}{2\sigma^2}\right),$$

with $\lambda > 0$, $\alpha > 0$, and $\beta > 0$.

Below we detail the solution step by step.

1.1 Likelihood $p(\mathbf{d}_{\text{obs}}|\mu, \sigma^2)$

Because every observation is drawn independently from the Gaussian

$$\mathcal{G}(d|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d - \mu)^2}{2\sigma^2}\right),$$

the joint likelihood is the product of these factors:

$$\begin{aligned} p(\mathbf{d}_{\text{obs}}|\mu, \sigma^2) &= \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d_{\text{obs},k} - \mu)^2}{2\sigma^2}\right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^{n} (d_{\text{obs},k} - \mu)^2\right). \end{aligned}$$

Thus, the likelihood is:

$$\boxed{p(\mathbf{d}_{\text{obs}}|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^{n} (d_{\text{obs},k} - \mu)^2\right].}$$

1.2 Conjugacy of the Gaussian–Inverse-Gamma Prior

Step 1: Writing the Joint Density

Our prior on (μ, σ^2) is

$$p(\mu, \sigma^2) = \frac{\sqrt{\lambda}}{\sqrt{2\pi\sigma^2}} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(-\frac{\lambda(\mu-\eta)^2 + 2\beta}{2\sigma^2}\right).$$

The joint posterior is proportional to the product of the likelihood and the prior:

$$\begin{aligned} p(\mu, \sigma^2|\mathbf{d}_{\text{obs}}) &\propto p(\mathbf{d}_{\text{obs}}|\mu, \sigma^2) p(\mu, \sigma^2) \\ &\propto (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(d_{\text{obs},k} - \mu)^2\right) \cdot \frac{1}{\sqrt{\sigma^2}} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(-\frac{\lambda(\mu-\eta)^2 + 2\beta}{2\sigma^2}\right) \\ &\propto (\sigma^2)^{-\left(\alpha+\frac{n}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}\left[\lambda(\mu-\eta)^2 + \sum_{k=1}^{n}(d_{\text{obs},k} - \mu)^2 + 2\beta\right]\right\}. \end{aligned}$$

Step 2: Completing the Square in μ

To show conjugacy, we must rearrange the exponent so that it appears as a quadratic form in μ. We start with

$$\lambda(\mu-\eta)^2 + \sum_{k=1}^{n}(d_{\text{obs},k} - \mu)^2.$$

It is convenient to introduce the sample mean:

$$\Phi^1_{\text{obs}} = \frac{1}{n}\sum_{k=1}^{n} d_{\text{obs},k}.$$

Also, note the following expansion:

$$\begin{aligned} \sum_{k=1}^{n}(d_{\text{obs},k} - \mu)^2 &= \sum_{k=1}^{n}\left[(d_{\text{obs},k} - \Phi^1_{\text{obs}}) + (\Phi^1_{\text{obs}} - \mu)\right]^2 \\ &= \sum_{k=1}^{n}(d_{\text{obs},k} - \Phi^1_{\text{obs}})^2 + n(\Phi^1_{\text{obs}} - \mu)^2, \end{aligned}$$

since the cross term vanishes (by definition of the sample mean).

Thus,

$$\lambda(\mu-\eta)^2 + \sum_{k=1}^{n}(d_{\text{obs},k} - \mu)^2 = \lambda(\mu-\eta)^2 + n(\mu - \Phi^1_{\text{obs}})^2 + \underbrace{\sum_{k=1}^{n}(d_{\text{obs},k} - \Phi^1_{\text{obs}})^2}_{\text{term independent of } \mu}.$$

We now complete the square with respect to μ in

$$\lambda(\mu-\eta)^2 + n(\mu - \Phi^1_{\text{obs}})^2.$$

Write:

$$\lambda(\mu-\eta)^2 + n(\mu - \Phi^1_{\text{obs}})^2 = (\lambda + n)\left(\mu - \eta'\right)^2 + C,$$

where

$$\eta' = \frac{\lambda\eta + n\Phi^1_{\text{obs}}}{\lambda + n},$$

and the constant C (which does not depend on μ) is

$$C = \frac{\lambda n}{\lambda + n}(\Phi^1_{\text{obs}} - \eta)^2.$$

Thus, the exponent in the posterior becomes

$$-\frac{1}{2\sigma^2}\left\{(\lambda + n)(\mu - \eta')^2 + \frac{\lambda n}{\lambda + n}(\Phi^1_{\text{obs}} - \eta)^2 + \sum_{k=1}^{n}(d_{\text{obs},k} - \Phi^1_{\text{obs}})^2 + 2\beta\right\}.$$

Step 3: Identifying the Updated Parameters

Now, grouping the terms, the joint posterior may be written (ignoring multiplicative factors not involving μ and σ^2) as

$$p(\mu, \sigma^2|\mathbf{d}_{\text{obs}}) \propto (\sigma^2)^{-\left(\alpha + \frac{n}{2} + 1\right)} \exp\left(-\frac{1}{2\sigma^2}\left[(\lambda + n)(\mu - \eta')^2 + 2\beta'\right]\right),$$

where we define

$$2\beta' = 2\beta + \frac{\lambda n}{\lambda + n}(\Phi^1_{\text{obs}} - \eta)^2 + \sum_{k=1}^{n}(d_{\text{obs},k} - \Phi^1_{\text{obs}})^2.$$

It is customary to write the sum of squares in terms of the unbiased sample variance estimator. Notice that

$$\Phi^2_{\text{obs}} = \frac{1}{n-1}\sum_{k=1}^{n}(d_{\text{obs},k} - \Phi^1_{\text{obs}})^2,$$

so that

$$\sum_{k=1}^{n}(d_{\text{obs},k} - \Phi^1_{\text{obs}})^2 = (n-1)\Phi^2_{\text{obs}}.$$

Thus, we identify the updated parameters of the Gaussian–inverse-Gamma posterior:

$$\begin{aligned}
\eta' &= \frac{\lambda\eta + n\Phi^1_{\text{obs}}}{\lambda + n}, \\
\lambda' &= \lambda + n, \\
\alpha' &= \alpha + \frac{n}{2}, \\
\beta' &= \beta + \frac{n\lambda}{2(\lambda + n)}(\Phi^1_{\text{obs}} - \eta)^2 + \frac{n-1}{2}\Phi^2_{\text{obs}}.
\end{aligned}$$

These updated parameters confirm that the posterior is also a Gaussian-inverse-Gamma distribution with parameters $(\eta', \lambda', \alpha', \beta')$.

Below is a detailed solution to the two questions.

2 Data compression

2.1 Sufficiency of the Statistics

Using the Neyman–Fisher Factorization Theorem

The factorization theorem states that a statistic $\mathbf{\Phi}(\mathbf{d})$ is sufficient for parameters (μ, σ^2) if we can write the likelihood $p(\mathbf{d}|\mu, \sigma^2)$ in the form

$$p(\mathbf{d}|\mu, \sigma^2) = g\Big(\mathbf{\Phi}(\mathbf{d}), \mu, \sigma^2\Big)\ h(\mathbf{d}),$$

where the function $h(\mathbf{d})$ does not depend on the parameters.

Step-by-Step Factorization

For n independent samples from $\mathcal{N}(\mu, \sigma^2)$, the likelihood is

$$p(\mathbf{d}|\mu, \sigma^2) = \Big(2\pi\sigma^2\Big)^{-\frac{n}{2}} \exp\Bigg\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(d_k - \mu)^2\Bigg\}.$$

A standard algebraic manipulation uses the identity

$$\sum_{k=1}^{n}(d_k - \mu)^2 = \underbrace{\sum_{k=1}^{n}(d_k - \Phi^1)^2}_{\text{data dispersion}} + n\Big(\Phi^1 - \mu\Big)^2.$$

Thus we can rewrite the likelihood as:

$$p(\mathbf{d}|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\Big\{-\frac{n}{2\sigma^2}(\Phi^1 - \mu)^2\Big\} \exp\Bigg\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(d_k - \Phi^1)^2\Bigg\}.$$

Notice here that: - The first exponential factor involves only the parameters μ, σ^2 and the empirical mean $\Phi^1(\mathbf{d})$. - The second exponential depends on the data only through $\sum_{k=1}^{n}(d_k - \Phi^1)^2$, which is exactly linked (by a multiplicative constant) to the sample variance $\Phi^2(\mathbf{d})$.

In more detail, since

$$\Phi^2(\mathbf{d}) = \frac{1}{n-1}\sum_{k=1}^{n}(d_k - \Phi^1)^2,$$

we have

$$\sum_{k=1}^{n}(d_k - \Phi^1)^2 = (n-1)\,\Phi^2(\mathbf{d}).$$

Hence, the likelihood can be written as:

$$p(\mathbf{d}|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\Big\{-\frac{n}{2\sigma^2}(\Phi^1 - \mu)^2\Big\} \exp\Big\{-\frac{(n-1)\,\Phi^2}{2\sigma^2}\Big\},$$

which now clearly factors into a function $g\Big(\Phi^1(\mathbf{d}), \Phi^2(\mathbf{d}), \mu, \sigma^2\Big)$ that depends on the parameters and the summary statistics and a function $h(\mathbf{d})$ (in this case, merely a constant with respect to μ, σ^2). Thus, by the factorization theorem, (Φ^1, Φ^2) is a sufficient statistic.

2.2 Distributions of $\Phi^1(\mathbf{d})$ and $\Phi^2(\mathbf{d})$

We now derive the closed forms for the distributions of the empirical mean and variance.

1. Distribution of Φ^1

Given that each d_k is independently distributed as $\mathcal{N}(\mu, \sigma^2)$ and using the standard result about sums of independent Gaussians, the sample mean

$$\Phi^1(\mathbf{d}) = \frac{1}{n}\sum_{k=1}^{n} d_k$$

follows a Gaussian distribution with mean μ and variance

$$\mathrm{Var}(\Phi^1) = \frac{\sigma^2}{n}.$$

That is,

$$\boxed{\Phi^1 \sim \mathcal{G}\left(\mu, \frac{\sigma^2}{n}\right).}$$

2. Distribution of Φ^2

Recall that the sample variance is given by

$$\Phi^2(\mathbf{d}) = \frac{1}{n-1}\sum_{k=1}^{n}(d_k - \Phi^1)^2.$$

A classical result in statistics is that if $d_k \sim \mathcal{N}(\mu, \sigma^2)$ independently, then

$$\frac{(n-1)\,\Phi^2}{\sigma^2} \sim \chi^2_{n-1}\,,$$

i.e. it is chi-squared distributed with $n-1$ degrees of freedom. A chi-squared distribution with k degrees of freedom is a special case of the Gamma distribution:

$$\chi^2_k \sim \mathit{\Gamma}\left(\frac{k}{2}, 2\right),$$

where the Gamma pdf is parametrised by a shape parameter (here, $k/2$) and a scale parameter (here, 2).

It then follows by a change of variable that

$$\Phi^2 \sim \mathit{\Gamma}\left(\frac{n-1}{2}, \frac{2\sigma^2}{n-1}\right).$$

That is, the pdf of Φ^2 is given by

$$p(\Phi^2) = \frac{1}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{n-1}{2\sigma^2}\right)^{\frac{n-1}{2}}(\Phi^2)^{\frac{n-1}{2}-1}\exp\left(-\frac{(n-1)\Phi^2}{2\sigma^2}\right).$$

Thus,

$$\boxed{\Phi^2 \sim \mathit{\Gamma}\left(\frac{n-1}{2}, \frac{2\sigma^2}{n-1}\right).}$$