Course Materials for GEN-AI

Northeastern University

These materials have been prepared and sourced for the course **GEN-AI** at Northeastern University. Every effort has been made to provide proper citations and credit for all referenced works.

If you believe any material has been inadequately cited or requires correction, please contact me at:

Instructor: Ramin Mohammadi
`r.mohammadi@northeastern.edu`

Thank you for your understanding and collaboration.

Normalization of Probability Distributions

Normalization of Probability Density Functions: Every probability density function (PDF) starts as an **unnormalized density** $f(z)$. To ensure it integrates to 1, it must be normalized by dividing by its **normalizing constant** $\mathcal{N}$, which is defined as:

$$\mathcal{N} = \int f(z) dz.$$

The normalized probability density $p(z)$ is then:

$$\boxed{p(z) = \frac{f(z)}{\mathcal{N}}},$$

which ensures:

$$\int p(z) dz = 1.$$

Examples of Normalizing Constants

1. Gaussian Distribution:

The unnormalized density is:

$$f(z) = \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right).$$

The normalizing constant is:

$$\mathcal{N} = \int_{-\infty}^{\infty} f(z) dz = \sqrt{2\pi\sigma^2}.$$

The normalized Gaussian becomes:

$$p(z) = \frac{f(z)}{\mathcal{N}} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right),$$

and it satisfies:

$$\int_{-\infty}^{\infty} p(z) dz = 1.$$

2. Uniform Distribution:

For a uniform distribution on $[a, b]$, the unnormalized density is:

$$f(z) = 1, \quad \text{for } z \in [a, b].$$

The normalizing constant is:

$$\mathcal{N} = \int_a^b f(z) dz = b - a.$$

The normalized density becomes:

$$p(z) = \frac{1}{b - a},$$

and it satisfies:

$$\int_a^b p(z) dz = 1.$$

Steps to Normalize $p(z)$

1. Before Normalization: Every $p(z)$ starts as an unnormalized density $f(z)$, and we compute:

$$\mathcal{N} = \int f(z) dz.$$

2. After Normalization: The normalized $p(z)$ is given by:

$$p(z) = \frac{f(z)}{\mathcal{N}},$$

and it satisfies:

$$\int p(z) dz = 1.$$

Why Is $\mathcal{N} = 1$ for $p(z)$ in AIS?

In Bayesian modeling and AIS, we typically work with distributions like Gaussian, Uniform, or Beta, which are already normalized by their definitions. For these distributions:

$$\mathcal{N} = 1 \quad \text{after normalization.}$$

For instance:

- Gaussian: $\mathcal{N} = \sqrt{2\pi\sigma^2}$ ensures $\int_{-\infty}^{\infty} p(z) dz = 1$.
- Uniform: $\mathcal{N} = b - a$ ensures $\int_a^b p(z) dz = 1$.

Since AIS assumes $p(z)$ is already normalized, the prior $p(z)$ satisfies:

$$\int p(z) dz = 1 \quad \text{and thus } \mathcal{N} = 1.$$

Expanded Understanding of Normalization

Normalization is a universal property of probability distributions. Here are additional clarifications:

- Every probability density function starts as an unnormalized function $f(z)$. The process of normalization ensures it can be treated as a proper distribution.
- In standard statistical applications, the normalizing constant $\mathcal{N}$ is precomputed for commonly used distributions, such as Gaussian, Uniform, Exponential, and Beta. This is why, in practice, we often treat distributions as "already normalized."
- In AIS and other inference techniques, we frequently rely on normalized distributions because the calculations of intermediate or final probabilities depend on this property.

Key Takeaways:

- Every density $f(z)$ must be normalized by dividing by its normalizing constant $\mathcal{N} = \int f(z)dz$.
- Standard distributions (e.g., Gaussian, Uniform) come pre-normalized, so $p(z)$ satisfies $\int p(z)dz = 1$ by default.
- In applications like AIS, the assumption $\mathcal{N}_1 = 1$ holds because $p(z)$ is already normalized.
- Normalization is essential for constructing valid probability distributions and for ensuring consistent calculations in Bayesian inference and machine learning applications.