Derivation of Gaussian Distribution

Haitian Xie

May 8, 2017

Suppose we have n random variables that are independent and identically distributed

$$X_1, X_2, \ldots, X_n \sim iid \tag{1}$$

which are observations of the true parameter μ (itself unobservable). Then the observation errors would be

$$X_1 - \mu, X_2 - \mu, \ldots, X_n - \mu \sim iid$$

We denote the density function of a single error by $f(\cdot)$, then the joint density of errors would be

$$F(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i - \mu)$$

which, from the Maximum Likelihood point of view, can also be seen as a likelihood function of the parameter μ

$$\mathcal{L}(\mu; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i - \mu) \tag{2}$$

where the observation values $x_1, x_2, \ldots, x_n$ are considered as given.

Now we make three basic assumptions for derivation:

1. $f(x) > 0, x \in \mathbb{R}$
2. $f(\cdot)$ is continuous differentiable
3. $\mathcal{L}(\mu)$ is maximized at $\mu = \bar{x} \coloneqq \frac{1}{n}\sum_{i=1}^{n} x_i$

The first two assumptions are only for the convenience of deduction, *i.e.* the ability to take log and derivative. The last assumption is really the essential one here. It suggests that the mean of the observations should be a good estimate of the true value μ, as it maximizes the likelihood function $\mathcal{L}(\mu)$. Then, because $log(\cdot)$ is a monotonically increasing function, we are able to perform logarithmic transformation on $\mathcal{L}$ without changing the maximum point $\mu = \bar{x}$. According to Fermat's theorem, $\mu = \bar{x}$ should be a stationary point of $log\left(\mathcal{L}(\mu)\right)$, which means

$$\left.\frac{\mathrm{d}log(\mathcal{L}(\mu))}{\mathrm{d}\mu}\right|_{\mu=\bar{x}} = 0 \tag{3}$$

From the chain rule and the computation rules of logarithm, we can deduct

$$\begin{aligned}
0 &= \frac{\mathrm{d}log\Big(\prod_{i=1}^{n} f(x_i-\mu)\Big)}{\mathrm{d}\mu}\bigg|_{\mu=\bar{x}} \\
&= \frac{\mathrm{d}\sum_{i=1}^{n} log(f(x_i-\mu))}{\mathrm{d}\mu}\bigg|_{\mu=\bar{x}} \\
&= \sum_{i=1}^{n} \frac{\mathrm{d}log(f(x_i-\mu))}{\mathrm{d}\mu}\bigg|_{\mu=\bar{x}} \\
&= \sum_{i=1}^{n} \frac{1}{f(x_i-\mu)}\frac{\mathrm{d}f(x_i-\mu)}{\mathrm{d}\mu}\bigg|_{\mu=\bar{x}} \\
&= \sum_{i=1}^{n} \frac{f'(x_i-\bar{x})}{f(x_i-\bar{x})}
\end{aligned}$$

Denote $g(\cdot) \coloneqq \frac{f'(\cdot)}{f(\cdot)}$, then $g(\cdot)$ is a continuous function. And now we have

$$\sum_{i=1}^{n} g(x_i-\bar{x}) = 0 \tag{4}$$

Before going further, we need to note here that equation (4) holds for any arbitrary values of $\vec{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $\forall n \in \mathbb{Z}^+$. The reason is that we assume $\bar{x}$ to be the maximum point of $\mathcal{L}(\mu)$ regardless of the number and values of observations $\vec{x}$. The only hidden constraint here is $\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

Consider the case $n = 2$, it is straightforward that

$$x_1 - \bar{x} = \frac{1}{2}(x_1 - x_2) = -\frac{1}{2}(x_2 - x_1) = x_2 - \bar{x}$$

together with the arbitrariness of the values of x_1 and x_2, equation (4) turns into

$$g(x) + g(-x) = 0, \forall x \in \mathbb{R} \tag{5}$$

which means $g(\cdot)$ is an odd function. Then consider the case $n = 3$, it is easy to see that

$$x_1 - \bar{x} = \frac{1}{3}(2x_1 - x_2 - x_3)$$

$$x_2 - \bar{x} = \frac{1}{3}(2x_2 - x_1 - x_3)$$

$$x_3 - \bar{x} = \frac{1}{3}(2x_3 - x_1 - x_2)$$

$$\implies -(x_3 - \bar{x}) = (x_1 - \bar{x}) + (x_2 - \bar{x}) \tag{6}$$

given the arbitrariness of x_1, x_2 and x_2, and that $g(\cdot)$ is odd, equation (4) turns into

$$g(x) + g(y) = g(x + y), \forall x, y \in \mathbb{R} \tag{7}$$

Equation (7) is **Cauchy's functional equation**. Given that $g(\cdot)$ is continuous, the equation has a single family of solutions

$$g(x) = ax, a \in \mathbb{R} \tag{8}$$

The proof is in the appendix.

By the definition of $g(\cdot)$, we have the following ordinary differential equation

$$\frac{f'(x)}{f(x)} = ax \tag{9}$$

$$\implies \Big(log(f(x))\Big)' = ax$$

$$\implies f(x) = c\exp{(\frac{a}{2}x^2)}, c > 0 \tag{10}$$

where c is the constant of integral. In order to be a density function, f needs to integrate to 1 on the real axis, therefore a must be negative for the (infinite limited) integral to converge, and also

$$1 = \int_{-\infty}^{\infty} f(x)dx = \sqrt{\frac{2\pi c^2}{|a|}}$$

$$\implies c = \sqrt{\frac{|a|}{2\pi}} \tag{11}$$

The computation of the integral is in the appendix. Now we have

$$f(x) = \sqrt{\frac{|a|}{2\pi}}\exp{(\frac{|a|}{2}x^2)}, a < 0$$

And we want to denote $\sigma^2 := \frac{1}{|a|}$. In that way,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp{(-\frac{x^2}{2\sigma^2})}, \forall \sigma^2 \in \mathbb{R} \tag{12}$$

and for any random variable X with density f, we have

$$var(X) = \int_{-\infty}^{\infty} x^2 f(x)dx = \sigma^2$$

Voilà, there is our Gaussian distribution.

Appendix

A. Suppose $g(\cdot)$ is a continuous function which satisfies

$$g(x) + g(y) = g(x + y), \forall x, y \in \mathbb{R}$$

prove that function g is scalar multiplication.

Proof: First, we want to show that the conclusion holds for any given rational number r.

(a) $r = 0$, obviously $g(0) = 0$

(b) $r > 0$, $\exists n, m \in \mathbb{Z}^+ \quad s.t. \quad r = \frac{n}{m}$

$$\begin{aligned} mg(r) = mg(\frac{n}{m}) &= g(\frac{n}{m}) + g(\frac{n}{m}) + \cdots + g(\frac{n}{m}) = g(\frac{n}{m} + \frac{n}{m} + \cdots + \frac{n}{m}) \\ &= g(n) = g(1) + g(1) + \cdots + g(1) = ng(1) \end{aligned}$$

$$\Longrightarrow \ g(r) = \frac{n}{m} g(1) = rg(1)$$

therefore we have $g(r) = g(1) \cdot r, \forall r \in \mathbb{Q}^+$

(c) $r < 0$, can be derived directly from the (b) and the fact that g is an odd function.

Then with the continuity of g and the fact that $\mathbb{Q}$ is dense in $\mathbb{R}$, we can derive the conclusion.

B. Calculate $\int_{-\infty}^{\infty} c \exp\left(\frac{a}{2}x^2\right) dx$ where $c > 0, a < 0$.

With the polar coordinates

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\frac{x^2+y^2}{2}) dxdy = \int_{0}^{2\pi}\int_{-\infty}^{\infty} \exp(-\frac{r^2}{2}) r dr d\theta = \int_{0}^{2\pi} d\theta = 2\pi$$

And the left hand side

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\frac{x^2+y^2}{2}) dxdy &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\frac{x^2}{2}) \exp(-\frac{y^2}{2}) dxdy \\ &= \Big(\int_{-\infty}^{\infty} \exp(-\frac{x^2}{2}) dx\Big)\Big(\int_{-\infty}^{\infty} \exp(-\frac{y^2}{2}) dy\Big) \\ &= \Big(\int_{-\infty}^{\infty} \exp(-\frac{x^2}{2}) dx\Big)^2 \end{aligned}$$

Therefore,

$$\int_{-\infty}^{\infty} \exp(-\frac{x^2}{2}) dx = \sqrt{2\pi}$$

Then with the substitution of $u = \sqrt{|a|}x$, we can easily derive that

$$\int_{-\infty}^{\infty} c \exp(\frac{a}{2}x^2) dx = c\sqrt{\frac{2\pi}{|a|}}$$

References

Cauchy's functional equation - Wikipedia
https://zhuanlan.zhihu.com/p/24437232