Stein's Method: a whirlwind tour

Lily Li

March 2021

1 Introduction

In these notes we summarize Cindy Zhang's survey of Nathan Ross' survey on the *Fundamentals of Stein's Method* with particular emphasis on the proof of the Central Limit Theorem. If you prefer to be talked at, consider Fraser Daly's video lectures. Throughout, let the Gaussian distribution be

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{1}$$

In 1972, Charles Stein, dissatisfied with the existing proof of the Central Limit Theorem, devised an alternative. We discuss his method and compare it to the standard Fourier Analysis approach. Central to this method is a characterization of the standard Gaussian random variable Z as one which satisfies: $\mathbb{E}f'(Z) = \mathbb{E}Zf(Z)$ for all well-behaved functions f (bounded and smooth). Any other random variable Y which approximately satisfies this identity is then close to Z in probability metric. We consider the Central Limit Theorem (CLT) as shown in Theorem 1.

Theorem 1. *(Central Limit Theorem). Let $X_1, ..., X_n$ be iid random variables with mean μ and variance σ^2. Define $S_n = X_1 + \cdots + X_n$ be their sum. Then*

$$\lim_{n\to\infty} \mathbb{P}\left[a < \frac{S_n - n\mu}{\sqrt{n\sigma^2}} < b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Recall its proof via Fourier Analysis. Use the notation from Theorem 1 and suppose for simplicity that $\mu = 0$ and $\sigma = 1$. Let the probability density function (pdf) of the X_is be $f(x)$. Then the pdf of S_n is the n-fold convolution of f i.e. $f^{*n}(x)$. Further the pdf of $\frac{S_n}{\sqrt{n}}$, the random variable that we would like to show can be approximated by $Z \sim \mathcal{N}(0, 1)$, has pdf $f_n(x) = \sqrt{n} f^{*n}(\sqrt{n}x)$ by a change of variables (see Lemma 9). Let $F(s) = \mathcal{F}f(s)$ be the Fourier Transform of f. Note that

$$\begin{aligned}
F\left(\frac{s}{\sqrt{n}}\right) &= \int_{-\infty}^{\infty} f(x)e^{-2\pi isx/\sqrt{n}}dx \\
&= \int_{-\infty}^{\infty} f(x)\left(1 - \frac{2\pi isx}{\sqrt{n}} - \frac{2\pi^2 s^2 x^2}{n} + o(1)\right)dx \\
&= \int_{-\infty}^{\infty} f(x)dx - \frac{2\pi isx}{\sqrt{n}}\int_{-\infty}^{\infty} xf(x)dx - \frac{2\pi^2 s^2}{n}\int_{-\infty}^{\infty} x^2 f(x)dx + \int_{-\infty}^{\infty} f(x)o(1)dx \\
&\approx 1 - \frac{2\pi^2 s^2}{n}
\end{aligned}$$

since f is the pdf of the X_is, $\mathbb{E}[X_i] = 0$, and $\mathbb{E}[X_i^2] = 1$. The Fourier Transform of $f_n(x)$ becomes

$$\left(\mathcal{F}\sqrt{n}f^{*n}(\sqrt{n}x)\right)(s) = \sqrt{n}F^n\left(\frac{s}{\sqrt{n}}\right) = \left(1 - \frac{2\pi^2 s^2}{n}\right)^n \approx e^{-2\pi^2 s^2} = \frac{1}{\sqrt{2\pi}}g\left(\frac{x}{2\pi}\right)$$

by making the change of variables $s = \frac{x}{2\pi}$ (see Lemma 10), which is exactly the Fourier Transform of the Gaussian distribution. Thus, by applying the inverse Fourier Transform and taking the limit as $n \to \infty$, the pdf $f_n(x)$ of $S_n/\sqrt{n}$ tends to $g(x)$.

To summarize, take the random variables $X_1, ..., X_n$ and analyze the pdf of their convolution. Since they are independent, the Fourier Transform of their convolution is the product of their Fourier Transforms . Observe that the Fourier Transform approximates the Gaussian so the original convolution must also approximate the Gaussian. Instead, Stein's approach characterizes the Gaussian and then shows that the normalized sum has approximately the same properties as the Gaussian.

2 Notation

In order to precisely capture *approximately the same*, we require the ability to compare distances between two probability distributions. Recall the definition of a **probability metric** for two probability measures μ and ν on a family $\mathcal{H}$ of test functions,

$$d_{\mathcal{H}}(\mu, \nu) = \sup_{h\in\mathcal{H}}\left|\int h(x)d\mu(x) - \int h(x)d\nu(x)\right|.$$

Typically the probability measures μ and ν are probability density functions for random variables Y and Z respectively. Then $d_{\mathcal{H}}(Y, Z) = \sup_{h\in\mathcal{H}}|\mathbb{E}h(Y) - \mathbb{E}h(Z)|$. By considering particular families, we can define particular metrics.

Definition 2. *(Wasserstein metric). Let W be the collection of 1-Lipschitz functions and Y, Z be two random variables. Then the — of Y and Z is*

$$d_W(Y, Z) = \sup_{h\in W}|\mathbb{E}h(Y) - \mathbb{E}h(Z)|.$$

Definition 3. *(Kolmogorov-Smirnov metric). Let K be the set of indicator function $\{\mathbb{1}_{\leq x} : x \in \mathbb{R}\}$ Y, Z be two random variables. Then the — of Y and Z is defined as*

$$d_K(Y, Z) = \sup_{x\in\mathbb{R}}|\mathbb{P}[Y \leq x] - P[Z \leq x]|.$$

In the case where $Z \sim \mathcal{N}(0,1)$ (and more generally when Z has bounded density in-terms of Lebesgue measure) we have the following bound on the the Kolmogorov metric in-terms of the Wasserstein metric.

Proposition 4. *(Bounding the Kolmogorov metric). For random variable Z with Lesbesgue density bounded by C and any other random variable Y,*

$$d_K(Y,Z) \leq \sqrt{2Cd_W(Y,Z)}.$$

Proof. Consider the indicator function $h_x(y) = \mathbb{1}_{y \leq x}$ and the "smoothed" function $h_{x,\epsilon}(y)$ which is 1 when $y \leq x$, 0 when $y > x + \epsilon$ and linear in-between. Then

$$\begin{aligned}
\mathbb{E}h_x(Y) - \mathbb{E}h_x(Z) &= \mathbb{E}h_x(Y) - \mathbb{E}h_{x,\epsilon}(Z) + \mathbb{E}h_{x,\epsilon}(Z) - \mathbb{E}h_x(Z) \\
&\leq \mathbb{E}h_{x,\epsilon}(Y) - \mathbb{E}h_{x,\epsilon}(Z) + \frac{C\epsilon}{2} \\
&\leq \frac{d_W(Y,Z)}{\epsilon} + \frac{C\epsilon}{2}
\end{aligned}$$

where the second line is because $\mathbb{E}h_x(Y) \leq \mathbb{E}h_{x,\epsilon}(Y)$ and the expectation of the amount that $h_{x,\epsilon}(Z)$ exceeds $\mathbb{E}h_x(Z)$ is exactly $C\epsilon/2$ (density of Z is bounded by C) and the third line is because $h_{x,\epsilon}$ is $(1/\epsilon)$-Lipschitz. By plugging in $\epsilon = \sqrt{2d_W(Y,Z)/C}$, we obtain the desired inequality. □

Note that C for $Z \sim \mathcal{N}(0,1)$ is $1/\sqrt{2\pi}$.

3 Characterizing the Gaussian

We characterize the Gaussian by the following identity.

Lemma 5. *(Stein's identity). If $Z \sim \mathcal{N}(0,1)$, then $\mathbb{E}f'(Z) = \mathbb{E}Zf(Z)$ for all absolute continuous functions $f : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}|f'(Z)| < \infty$.*

Conversely, if $\mathbb{E}f'(Z) = \mathbb{E}Zf(Z)$ for all bounded, continuous, and piece-wise continuously differentiable functions f with $\mathbb{E}|f'(Z)| < \infty$, then $Z \sim \mathcal{N}(0,1)$.

Proof. The forward direction follows from the definition of $Z \sim \mathcal{N}(0,1)$. In particular,

$$\begin{aligned}
\mathbb{E}Zf(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} zf(z)e^{-z^2/2}dz \\
&= \frac{-f(z)}{\sqrt{2\pi}} e^{-z^2/2}\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(z)e^{-z^2/2}dz \\
&= \mathbb{E}f'(Z)
\end{aligned}$$

where we integrate-by-parts on the second line and the third line follows since $\mathbb{E}[|f'(Z)|]$ is bounded.

In the reverse direction, we suppose that $\mathbb{E}f'(Z) = \mathbb{E}Zf(Z)$ for all bounded, continuous, and piece-wise continuously differentiable functions f. Let $\Phi(z) = \mathbb{P}(Z \leq z)$ be the cumulative distribution function (cdf) of Z and fix some $z \in \mathbb{R}$. Pay particular attention to

$$f(y) = \sqrt{2\pi}e^{y^2/2} \min_{x \in \{y,z\}} \Phi(x) \left(1 - \max_{x' \in \{y,z\}} \Phi(x')\right). \tag{2}$$

We will show that f exactly satisfies Stein's Equation:

$$f'(y) - yf(y) = \mathbb{1}_{y \leq z} - \Phi(z).$$

Since f is a bounded, continuous, and piece-wise continuously differentiable function, the expectation of the LHS is zero by assumption. It follows that the expectation of the RHS is zero as well so the cdf of Y approaches the cdf of the standard normal in the limit.

To see that f satisfies the above equation, solve the ODE for f as follows.

$$\begin{aligned} \left(f(y)e^{-y^2/2}\right)' &= e^{-y^2/2}\left(f'(y) - yf(y)\right) = e^{-y^2/2}\left(\mathbb{1}_{y \leq z} - \Phi(z)\right) \\ f(y) &= e^{y^2/2}\int_{-\infty}^{y} e^{-x^2/2}\left(\mathbb{1}_{x \leq z} - \Phi(z)\right)dx \\ &= -e^{y^2/2}\int_{y}^{\infty} e^{-x^2/2}\left(\mathbb{1}_{x \leq z} - \Phi(z)\right)dx. \end{aligned}$$

Using the last identity in the case where $y > z$, we obtain

$$f(y) = e^{y^2/2}\int_{y}^{\infty} e^{-x^2/2}\Phi(z)dx = \sqrt{2\pi}e^{y^2/2}\Phi(z)\left(1 - \Phi(y)\right).$$

Using the penultimate identity when $y \leq z$ obtains the other result. □

From this identity, if some random variable Y satisfies $\mathbb{E}f'(Y) \approx \mathbb{E}Yf(Y)$, then we would expect Y to be approximately Gaussian. More precisely, we would like to bound the probability metric of Y and Z as

$$d_{\mathcal{H}}(Y,Z) = \sup_{h \in \mathcal{H}} |\mathbb{E}h(Y) - \mathbb{E}h(Z)| \leq \sup_{f \in \mathcal{F}} |\mathbb{E}f'(Y) - Y\mathbb{E}f(Y)| \tag{3}$$

where the RHS tends to zero. To this end, define Stein's equation:

$$f'(y) - yf(y) = h(y) - \mathbb{E}h(Z). \tag{4}$$

By taking expectations then the sup of both sides, we can recover the inequality of Equation 3. In a typical Stein's method use case, the function h is given by the probability metric and we want to find the f which satisfies Stein's equation.

4 Application: Proof of CLT

We can prove CLT using the above. See Ross [3] for details. Let $X_1, ..., X_n$ be independent mean zero random variables with $\mathbb{E}|X_i|^4$ and $\mathbb{E}X_i^2 = 1$. If $Y = \left(\sum X_i\right)/\sqrt{n}$ and $Z \sim \mathcal{N}(0,1)$, then

$$d_W(Y,Z) \leq \sup_{f \in \mathcal{F}} \left|\mathbb{E}\left[f'(Y) - Yf(Y)\right]\right| \leq \sqrt{\frac{2}{\pi n^2}\sum \mathbb{E}|X_i|^4} + \frac{1}{n^{3/2}}\sum \mathbb{E}|X_i|^3.$$

5 Variant: Method of Exchangeable Pairs

Instead we will show a proof of CLT via the method of exchangeable pairs. First, for an ordered pair (Y, Y') of random variables , it is an **exchangeable pair**[1] if $(Y, Y') =^d (Y', Y)$. For some $0 < \lambda \leq 1$, an exchangeable pair (Y, Y') which satisfies $\mathbb{E}[Y'|Y] = (1-\lambda)Y$ is an λ-Stein pair. We note the following properties of an λ-Stein pair.

Proposition 6. *(Properties of λ-Stein Pairs). Let $F : \mathbb{R}^2 \to \mathbb{R}$ be an anti-symmetric function*[2] *and (Y, Y') be an λ-Stein pair with $\operatorname{Var}(Y) = \sigma^2$, then*

1. $\mathbb{E}F(Y, Y') = 0$.
2. $\mathbb{E}Y = 0$ *and* $\mathbb{E}(Y' - Y)^2 = 2\lambda\sigma^2$.

Proof. The first item follows by exchangeability and anti-symmetry: $\mathbb{E}F(Y, Y') = \mathbb{E}F(Y', Y) = -\mathbb{E}F(Y, Y')$. The first equality of the second item follows by conditional expectations where

$$\mathbb{E}Y = \mathbb{E}Y' = \mathbb{E}\mathbb{E}[Y'|Y] = (1-\lambda)\mathbb{E}Y.$$

The second, also by conditional expectation and by explicitly computing $\mathbb{E}(Y' - Y)^2$.

$$\mathbb{E}(Y' - Y)^2 = \mathbb{E}\left[(Y')^2 + Y^2 - 2\mathbb{E}\left[Y'|Y\right]Y\right] = 2\sigma^2 - 2(1-\lambda)\sigma^2 = 2\lambda\sigma^2$$

since Y and Y' have the same distribution. □

Note: the first property holds for general exchangeable pairs. From the second property we can get a sense of the effect of λ on the Stein pair: the smaller the λ, the smaller the variance of $Y' - Y$.

If we can find a Y' such that (Y, Y') is an λ-Stein pair, then we can use it to bound $d_W(Y, Z)$.

Theorem 7. *(Bounding Wasserstein metric using λ-Stein pairs). If (Y, Y') is an λ-Stein pair with $\mathbb{E}Y^2 = 1$ and $Z \sim \mathcal{N}(0, 1)$, then*

$$d_W(Y, Z) = \sqrt{\frac{\operatorname{Var}\left(\mathbb{E}[(Y' - Y)^2|Y]\right)}{2\pi\lambda^2}} + \frac{1}{3\lambda}\mathbb{E}|Y' - Y|^3.$$

Proof. We want to rewrite $\mathbb{E}Yf(Y)$ in such a way so that it is evidently close to $\mathbb{E}f'(Y)$. To this end, suppose that f is bounded with bounded first derivative (by $\sqrt{2/\pi}$) and second derivative (by 2). Let $F(y) := \int_0^y f(t)dt$. By exchangeability, we have $0 = \mathbb{E}[F(Y') - F(Y)]$. By Taylor's expansion we can rewrite this as:

$$0 = \mathbb{E}\left[(Y' - Y)f(Y) + \frac{(Y' - Y)^2 f'(Y)}{2} + \frac{(Y' - Y)^3 f''(Y^*)}{6}\right]$$

[1] Being exchangeable pairs *imples* same distribution, but the converse is false. In-terms of the matrix of marginals (which is square since the support of the two random variables are the same), Y and Y' are exchangeable implies that the matrix is symmetric while Y and Y' have the same distribution only implies that the corresponding row and column sums are equal.

[2] F is **anti-symmetric** if $F(x, y) = -F(y, x)$.

where Y^* is some quantity in the interval with endpoints Y and Y'. Further, by the Stein pair condition we have:

$$\mathbb{E}[(Y'-Y)f(Y)] = \mathbb{E}[f(Y)\mathbb{E}[(Y'-Y)|Y]] = -\lambda\mathbb{E}[Yf(Y)].$$

Combining the above, we have

$$\begin{aligned}\mathbb{E}[Yf(Y)] &= \mathbb{E}\left[\frac{(Y'-Y)^2f'(Y)}{2a} + \frac{(Y'-Y)^3f''(Y^*)}{6a}\right] \text{ and}\\ \mathbb{E}[f'(Y)-Yf(Y)] &\leq \|f'\|\left|1-\frac{\mathbb{E}[(Y'-Y)^2|Y]}{2\lambda}\right| + \|f''\|\frac{\mathbb{E}|Y'-Y|^3}{6\lambda}\\ &\leq \sqrt{\frac{\mathrm{Var}\left(\mathbb{E}[(Y'-Y)^2|Y]\right)}{2\pi\lambda^2}} + \frac{1}{3\lambda}\mathbb{E}|Y'-Y|^3\end{aligned}$$

where we get the last inequality by applying Cauchy-Schwarz and plugging in all the bounds. □

Next, to prove the central limit theorem, let $X_1,...,X_n$ be independent with $\mathbb{E}X_i^4 < \infty$, $\mathbb{E}X_i = 0$, $\mathrm{Var}(X_i) = 1$, and $Y = n^{-1/2}\sum_{i=1}^n X_i$. Construct an exchangeable pair by picking an index uniformly at random and replacing it with an independent copy i.e. let I be uniform on $\{1,...,n\}$ and $(X_1',...,X_n')$ an independent copy of $(X_1,...,X_n)$, then define

$$Y' = Y - \frac{X_I}{\sqrt{n}} + \frac{X_I'}{\sqrt{n}}.$$

We show that (Y,Y') forms a $(1/n)$-Stein pair.

$$\mathbb{E}[Y'-Y|Y] = \frac{1}{\sqrt{n}}\mathbb{E}[X_I'-X_I|Y] = \frac{1}{n}\sum_{i=1}^n\frac{\mathbb{E}[X_i'-X_i|Y]}{\sqrt{n}} = -\frac{Y}{n}.$$

Apply Theorem 7 and bound $\frac{1}{\sqrt{2\pi}\lambda}\mathrm{Var}\left(\mathbb{E}\left[(Y'-Y)^2|Y\right]\right)$ and $\frac{1}{3\lambda}\mathbb{E}|Y'-Y|^3$. For the first term,

$$\mathbb{E}[(Y'-Y)^2|Y] = \frac{1}{n}\mathbb{E}[(X_I-X_I')^2|X_I] = \frac{1}{n^2}\sum_{i=1}^n\mathbb{E}[(X_i-X_i')^2|X_i] = \frac{1}{n^2}\sum_{i=1}^n\left(1+X_i^2\right)$$

With the variance and adding in the coefficients,

$$\frac{1}{\sqrt{2\pi}\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[(Y'-Y)^2|Y\right]\right)} \leq \frac{n}{\sqrt{2\pi}}\sqrt{\frac{1}{n^4}\sum_{i=1}^n\mathbb{E}X_i^4} = \sqrt{\frac{1}{2\pi n^2}\sum_{i=1}^n\mathbb{E}X_i^4}$$

Further, for the second term in Theorem 7, we have

$$\frac{1}{3\lambda}\mathbb{E}|Y'-Y|^3 = \frac{n}{3n^{3/2}}\mathbb{E}|X_I-X_I'|^3 = \frac{1}{3n^{3/2}}\sum_{i=1}^n\mathbb{E}|X_i-X_i'|^3 \leq \frac{8}{3n^{3/2}}\sum_{i=1}^n\mathbb{E}|X_i|^3$$

Together, the Wasserstein metric — and thus the Kolmogorov metric — is bounded by

$$d_W(Y,Z) \leq \sqrt{\frac{1}{2\pi n^2}\sum_{i=1}^n\mathbb{E}X_i^4} + \frac{8}{3n^{3/2}}\sum_{i=1}^n\mathbb{E}|X_i|^3.$$

The RHS tends to zero at a rate of $n^{-1/2}$ since $\mathbb{E}X_i^4 < \infty$ and thus $\mathbb{E}|X_i|^3 < \infty$.

6 Conclusion

The same method can be used to show convergence to other distributions: Poisson, multi-variate normal, gamma, beta, etc. See this talk at MIT by Gesine Reinert. The standard monograph for this topic is the one by Diaconis and Holmes [2]. The standard textbook is the one by Barbour and Chen [1].

References

[1] Andrew D Barbour and Louis Hsiao Yun Chen. *An introduction to Stein's method*, volume 4. World Scientific, 2005.

[2] Persi Diaconis and Susan Holmes. Stein's method: expository lectures and applications. IMS, 2004.

[3] Nathan Ross et al. Fundamentals of steins method. *Probability Surveys*, 8:210–293, 2011.

A Appendix

The following are some useful properties of the Gaussian distribution and Fourier Transforms.

Lemma 8. *(Gaussian pdf is normalized). For $g(x)$ from Equation 1, $\int_{-\infty}^{\infty} g(x)dx = 1$.*

Proof. Instead of integrating $g(x)$ directly, we will instead consider

$$I = \int_{-\infty}^{\infty} e^{-ax^2} dx$$

for any constant a. In particular we compute I^2 using polar coordinates.

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} e^{-ax^2} dx \right) \cdot \left(\int_{-\infty}^{\infty} e^{-ay^2} dy \right) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-a(x^2+y^2)} dxdy \\ &= \int_{0}^{2\pi} \int_{0}^{\infty} re^{-ar^2} drd\theta = \int_{0}^{2\pi} \frac{d\theta}{2a} = \frac{\pi}{a}. \end{aligned}$$

Note that $dxdy = rdrd\theta$ since we take into account the determinant of the Jacobian when transforming from the (x, y) regime into the (r, θ) regime. By plugging in $a = 1/2$, we see that the normalization factor of I is exactly $\frac{1}{\sqrt{2\pi}}$. □

Lemma 9. *(Fourier Transform under similarity). For any function f in $L^1(\mathbb{R})$ and constant $a > 0$, $f(x)$ is to $\mathcal{F}f(s)$ as $f(ax)$ is to $(1/a)\mathcal{F}f(s/a)$.*

Proof. Make a change of variables $u = ax$ in the definition of the Fourier Transform to obtain:

$$
\begin{aligned}
\left(\mathcal{F}f(ax)\right)(s) &= \int_{-\infty}^{\infty} f(ax)e^{-2\pi isx}dx \\
&= \frac{1}{a}\int_{-\infty}^{\infty} f(u)e^{-2\pi i\frac{s}{a}u}du \\
&= \frac{1}{a}\mathcal{F}f\left(\frac{s}{a}\right)
\end{aligned}
$$

□

Lemma 10. *(Fourier Transform of the Gaussian). Let $g(x)$ be as shown in Equation 1. Then $\mathcal{F}g(s) = \frac{1}{\sqrt{2\pi}}g\left(\frac{s}{2\pi}\right)$.*

Proof. By definition, the Fourier Transform of $g(x)$ is

$$
\mathcal{F}g(s) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}e^{-2\pi ixs}dx
$$

Applying derivatives to both sides, we see that

$$
\begin{aligned}
\frac{d\mathcal{F}g(s)}{ds} &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\frac{de^{-2\pi ixs}}{ds}dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -2\pi ixe^{-x^2/2}e^{-2\pi ixs}dx.
\end{aligned}
$$

Integrating by parts where $\int u'v = uv - \int uv'$ with $u = e^{-x^2/2}$ and $v = 2\pi ie^{-2\pi ixs}$, we see that

$$
\begin{aligned}
\mathcal{F}'g(s) &= \frac{1}{\sqrt{2\pi}}\left(2\pi ie^{-x^2/2}e^{2\pi isx}\big|_{x\in\mathbb{R}} - \int_{-\infty}^{\infty} 4\pi^2 se^{-x^2/2}e^{-2\pi ixs}dx\right) \\
&= -4\pi^2 s\mathcal{F}g(s).
\end{aligned}
$$

Solving this ODE, we see that $\mathcal{F}g(s) = e^{-2\pi^2s^2}\mathcal{F}g(0) = e^{-2\pi^2s^2} = \frac{1}{\sqrt{2\pi}}g\left(\frac{s}{2\pi}\right)$. □