The Student t Distribution

Study Notes — March 19, 2011 — Kenneth Chu (300517641)

1 Motivation

Let $\{ X_i : \Omega \longrightarrow \mathbb{R} \}_{i \in \mathbb{N}}$ be a sequence of independent and identically distributed (*i.i.d.*) random variables, with common mean (expected value) $\mu := E(X_i)$, for any $i \in \mathbb{N}$, and finite variance $\sigma^2 > 0$. For each $n \in \mathbb{N}$, let

$$\overline{X}_n \;:=\; \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{and} \qquad Z_n \;:=\; \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

The random variable $\overline{X}_n$ is called the *sample mean* (of the sample consisting of $X_1, \ldots, X_n$). Then,

- $E\big(\overline{X}_n\big) = \mu$ and $\mathrm{Var}\big(\overline{X}_n\big) = \dfrac{\sigma^2}{n}$. Thus, $\overline{X}_n$ is an unbiased estimator for the parameter μ, and for large n, any observed value of $\overline{X}_n$ is expected to closely approximate μ (since $\mathrm{Var}\big(\overline{X}_n\big) \longrightarrow 0$ as $n \longrightarrow \infty$).

  Thus, we may estimate the value of the parameter μ by taking the average $\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$ of a set of sampled values $x_1, \ldots, x_n$ of $X_1, \ldots, X_n$, respectively.

- Assume the value of $\sigma^2 > 0$ is known while that of μ is not known. Then, given any explicit candidate value μ_0 for the unknown parameter μ, we may assess the reliability of the hypothesis $\mu = \mu_0$ by taking sampled values $x_1, \ldots, x_n$ for $X_1, \ldots, X_n$, respectively.

  Given the sampled values $x_1, \ldots, x_n$, define $\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$. By the Central Limit Theorem, the distribution of Z_n approaches the standard normal distribution $\mathcal{N}(0,1)$ as $n \longrightarrow \infty$, which allows us to approximate the conditional probability $P\big(\,\big|\overline{X}_n - \mu\big| \geq |\overline{x} - \mu| \;\big|\; \mu = \mu_0\big)$ as follows:

  $$\begin{aligned} P\big(\,\big|\overline{X}_n - \mu\big| \geq |\overline{x} - \mu| \;\big|\; \mu = \mu_0\big) &= P\left(\frac{\big|\overline{X}_n - \mu\big|}{\sigma/\sqrt{n}} \geq \frac{|\overline{x} - \mu|}{\sigma/\sqrt{n}} \;\middle|\; \mu = \mu_0 \right) \\ &\approx P\left(|\,Z_n\,| \geq \frac{|\overline{x} - \mu|}{\sigma/\sqrt{n}} \;\middle|\; \mu = \mu_0 \right), \quad \text{for large } n. \end{aligned}$$

  The conditional probability $P\left(|\,Z\,| \geq \dfrac{|\overline{x} - \mu|}{\sigma/\sqrt{n}} \;\middle|\; \mu = \mu_0 \right)$ is called the p-value corresponding to the set $\{x_1, \ldots, x_n\}$ of sampled values under the null hypothesis $\mu = \mu_0$. If the p-value subceeds a certain threshold α (common values for α include 0.05 and 0.01), the null hypothesis $\mu = \mu_0$ is rejected, the underlying intuition being that, under the assumption $\mu = \mu_0$, the probability of obtaining a set of sampled values "as extreme as or more extreme than" $\{x_1, \ldots, x_n\}$ is "too low" (i.e. subceeding α).

However, the hypothesis test mentioned above relies on the requirement that the value of $\sigma^2 > 0$ be known. In practice, this is seldom the case. In the predominant case that the value of σ^2 is unknown, we could only approximate the value of σ^2 based on sampled data somehow.

To this end, we now assume that $X_1, X_2, \ldots \sim \mathcal{N}(\mu, \sigma^2)$, i.e. the random variables $X_1, X_2, \ldots$ are *i.i.d.* normal random variables, where the values of both μ and σ^2 are not known. The maximum likelihood estimators for μ and σ^2, based on sampled values for $X_1, \ldots, X_n$, are then respectively

$$\widehat{\mu}_{\mathrm{MLE}} \;=\; \overline{X}_n \;:=\; \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{and} \qquad \widehat{\sigma^2}_{\mathrm{MLE}} \;=\; \frac{1}{n}\sum_{i=1}^{n}\big(X_i - \overline{X}_n\big)^2.$$

See Example 4, §8.3, [1] for the derivation of $\widehat{\mu}_{\mathrm{MLE}}$ and $\widehat{\sigma^2}_{\mathrm{MLE}}$. Now, $E(\widehat{\mu}_{\mathrm{MLE}}) = E\big(\overline{X}_n\big) = \mu$; so, $\widehat{\mu}_{\mathrm{MLE}}$ is an unbiased estimator for μ. On the other hand,

$$E\left(\sum_{i=1}^{n}\big(X_i - \overline{X}_n\big)^2 \right) \;=\; (n-1)\,\sigma^2.$$

Consequently,

$$E\Big(\widehat{\sigma^2}_{\mathrm{MLE}}\Big) \;=\; E\left(\frac{1}{n}\sum_{i=1}^{n}\big(X_i - \overline{X}_n\big)^2 \right) \;=\; \left(\frac{n-1}{n}\right)\sigma^2.$$

Thus, $\widehat{\sigma^2}_{\mathrm{MLE}}$ is NOT an unbiased estimator for σ^2, but S_n^2 is an unbiased estimator for σ^2, where S_n^2, called the *unbiased sample variance*, is defined by:

$$S_n^2 \;:=\; \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2.$$

Hence, in the case that $X_1, X_2, \ldots$ are *i.i.d.* normal random variables with common distribution $\mathcal{N}(\mu, \sigma^2)$, we may assess the reliability of the hypothesis $\mu = \mu_0$, for some candidate value μ_0 for the parameter μ, provided we know the probability distribution of ***Student's t ratio***, which is defined by:

$$T_{n-1} \;:=\; \frac{\overline{X}_n - \mu}{\sqrt{S_n^2 \,/\, n}}.$$

The probability distribution of T_{n-1} is the ***Student t distribution with $(n-1)$ degrees of freedom***.

2 Summary

- Let $X_1, X_2, \ldots, X_n$ be *i.i.d.* normal random variables with common mean μ and finite variance $\sigma^2 > 0$.

$$\overline{X}_n \;:=\; \frac{1}{n}\sum_{i=1}^{n} X_i\,, \qquad \text{and} \qquad S_n^2 \;:=\; \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2.$$

  $\overline{X}_n$ is called the ***sample mean***, and S_n^2 is called the (***unbiased***) ***sample variance***. They are random variables and are unbiased estimators for μ and σ^2, respectively.

- Note that

$$T_{n-1} \;:=\; \frac{\overline{X}_n - \mu}{\sqrt{S_n^2 \,/\, n}} \;=\; \frac{\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\left(\dfrac{(n-1)\,S_n^2}{\sigma^2}\right)\Big/\,(n-1)}}.$$

  We claim:

  - The numerator $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ of T_{n-1} has the standard normal distribution.
  - The term $\dfrac{(n-1)\,S_n^2}{\sigma^2}$ in the denominator of T_{n-1} is a random variable whose distribution is a χ^2 distribution with $(n-1)$ degrees of freedom.
  - The two random variables $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ and $\dfrac{(n-1)\,S_n^2}{\sigma^2}$ are independent of each other.

- **Definition 2.1**
  Let Z be a standard normal random variable and X be a χ^2 random variable with n degrees of freedom. Suppse Z and X are independent. The ***Student t distribution with n degrees of freedom*** *is the probability distribution of the following random variable:*

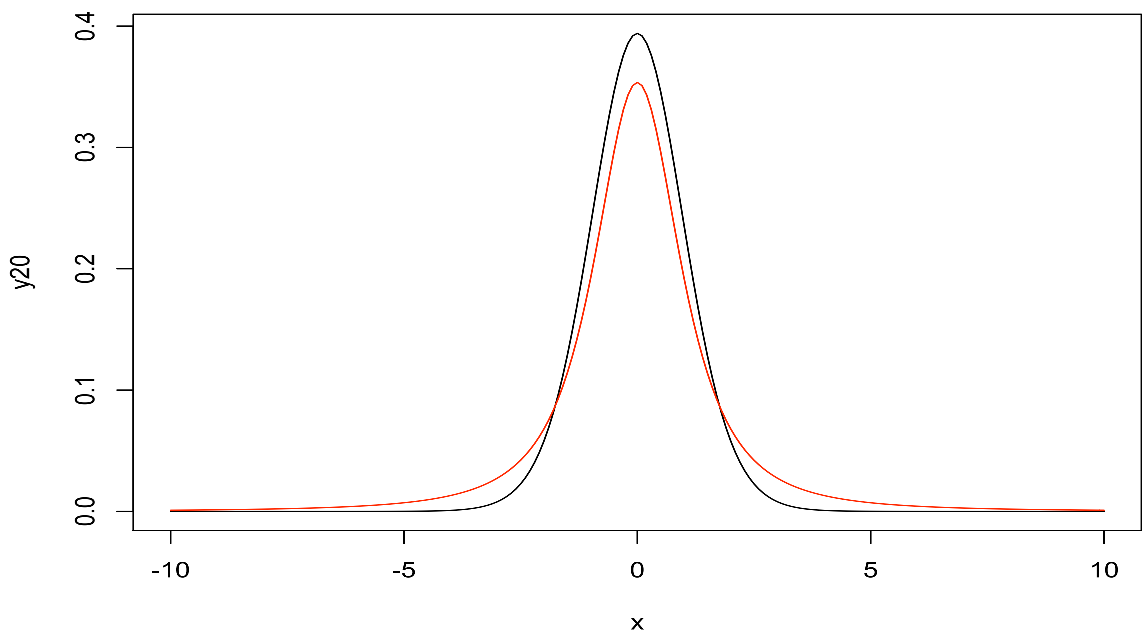
$$T_n \;:=\; \frac{Z}{\sqrt{X/n}}.$$

  The random variable T_n is called the ***Student's t ratio with n degrees of freedom***.

- **Theorem 2.2**
  The probability density function of the Student t distribution with n degrees of freedom is given by:

$$f_{T_n}(t) \;=\; \frac{\Gamma\!\left(\dfrac{n+1}{2}\right)}{\sqrt{n\,\pi}\,\Gamma\!\left(\dfrac{n}{2}\right)} \cdot \frac{1}{\left(1 + \dfrac{t^2}{n}\right)^{(n+1)/2}}\,, \quad \text{for} \quad -\infty < t < \infty.$$

The diagram below shows the graphs of the probability density functions $f_{T_{20}}$ in black and f_{T_2} in red.



The above graph is generated with `R` with the following command:

```
> y20 = dt(x,df=20); y2 = dt(x,df=2); plot(x,y20,type="l"); points(x,y2,type="l",col="red");
```

3 χ^2_n — the distribution of the sum of squares of $n \in \mathbb{N}$ independent standard normal random variables

Theorem 3.1

Let $Z_1, Z_2, \ldots, Z_n$ be n independent standard normal random variables, and let $X := \sum_{i=1}^{n} Z_i^2$. Then, X has a Gamma distribution with parameter values $r = \dfrac{n}{2}$ and $\lambda = \dfrac{1}{2}$. Equivalently, the probability density function of X is given by

$$f_X(x) \;=\; \frac{1}{2^{n/2}\,\Gamma\!\left(\dfrac{n}{2}\right)}\, x^{(n/2)-1}\, e^{-x/2}\,, \quad \text{for } x \geq 0.$$

PROOF First, take $m = 1$. We claim that $Z^2 \sim \Gamma(r, \lambda)$, for $r = \lambda = \dfrac{1}{2}$, where $\Gamma(r, \lambda)$ denotes the Gamma distribution. Indeed, for any $x \geq 0$,

$$\begin{aligned}
F_{Z^2}(x) &= P\big(Z^2 \leq x\big) \;=\; P\big(-\sqrt{x} \leq Z \leq \sqrt{x}\big) \;=\; 2\,P\big(0 \leq Z \leq \sqrt{x}\big) \\
&= 2\int_0^{\sqrt{x}} f_Z(\zeta)\,\mathrm{d}\zeta \;=\; 2\int_0^{\sqrt{x}} \frac{1}{\sqrt{2\,\pi}}\, e^{-\zeta^2/2}\,\mathrm{d}\zeta \\
&= \sqrt{\frac{2}{\pi}} \int_0^{\sqrt{x}} e^{-\zeta^2/2}\,\mathrm{d}\zeta.
\end{aligned}$$

Differentiating $F_{Z^2}(x)$ with respect to x yields:

$$\begin{aligned} f_{Z^2}(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_{Z^2}(x) = \sqrt{\frac{2}{\pi}} \frac{\mathrm{d}}{\mathrm{d}x}\left(\int_0^{\sqrt{x}} e^{-\zeta^2/2}\,\mathrm{d}\zeta\right) = \sqrt{\frac{2}{\pi}}\, e^{-x/2} \frac{\mathrm{d}}{\mathrm{d}x}\big(\sqrt{x}\big) = \sqrt{\frac{2}{\pi}}\, e^{-x/2} \frac{1}{2\sqrt{x}} \\ &= \frac{1}{\sqrt{2}\,\Gamma\big(\frac{1}{2}\big)}\, x^{(1/2)-1}\, e^{-x/2}, \quad \text{for } x > 0. \end{aligned}$$

We have used the fact that $\Gamma\big(\frac{1}{2}\big) = \sqrt{\pi}$. Thus, $f_{Z^2}(x)$ is the probability density function of the Gamma distribution $\Gamma(r,\lambda)$ with $r = \dfrac{1}{2}$ and $\lambda = \dfrac{1}{2}$, since for any $r, \lambda > 0$,

$$f_{\Gamma(r,\lambda)}(y) = \frac{\lambda^r}{\Gamma(r)}\, y^{r-1}\, e^{-\lambda y}, \quad \text{for } y \geq 0.$$

Next, recall that: If $G_1 \sim \Gamma(r_1,\lambda)$ and $G_2 \sim \Gamma(r_2,\lambda)$ are independent random variables, then $G_1 + G_2 \sim \Gamma(r_1 + r_2, \lambda)$. Induction thus immediately gives: If $G_i \sim \Gamma(r_i,\lambda)$, $i = 1,\ldots,n$, are independent random variables, then

$$\sum_{i=1}^{n} G_i \sim \Gamma\left(\sum_{i=1}^{n} r_i\,,\,\lambda\right).$$

(See, for example, Theorem 4.6.4, [3], or §2.4, [4]).
Since $Z_1^2, Z_2^2, \ldots, Z_n^2 \sim \Gamma\big(r = \frac{1}{2}, \lambda = \frac{1}{2}\big)$ and they are independent random variables, it now follows that

$$X := \sum_{i=1}^{n} Z_i^2 \sim \Gamma\left(r = \frac{n}{2}\,,\,\lambda = \frac{1}{2}\right).$$

In other words, the probability density function of X is given by:

$$f_X(x) = f_{\Gamma(r=n/2,\lambda=1/2)}(x) = \frac{\big(\frac{1}{2}\big)^{n/2}}{\Gamma\big(\frac{n}{2}\big)}\, x^{(n/2)-1}\, e^{-(1/2)\cdot x} = \frac{1}{2^{n/2}\,\Gamma\left(\dfrac{n}{2}\right)}\, x^{(n/2)-1}\, e^{-x/2}, \quad \text{for } x > 0,$$

as required. □

Definition 3.2
The ***chi-square distribution with*** $n \in \mathbb{N}$ ***degree(s) of freedom*** *is, by definition, Gamma distribution* $\Gamma\big(r = \frac{n}{2}, \lambda = \frac{1}{2}\big)$. *It is denoted by* χ_n^2.

Remark 3.3
The sum $X := Z_1^2 + Z_2^2 + \cdots + Z_n^2$ *of the squares of* $n \in \mathbb{N}$ *independent standard normal random variables* $Z_1, Z_2, \ldots, Z_n \sim \mathcal{N}(0,1)$ *has a chi-square distribution with* n *degree(s) of freedom. In other words,* $X := Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi_n^2$.

Remark 3.4
Note that

$$f_{\chi_1^2}(x) = \frac{1}{2^{1/2}\,\Gamma\left(\dfrac{1}{2}\right)}\, x^{(1/2)-1}\, e^{-x/2} = \frac{1}{\sqrt{2}\,\sqrt{\pi}}\, x^{-1/2}\, e^{-x/2} = \frac{1}{\sqrt{2\pi}\cdot x^{1/2}\cdot e^{x/2}}$$

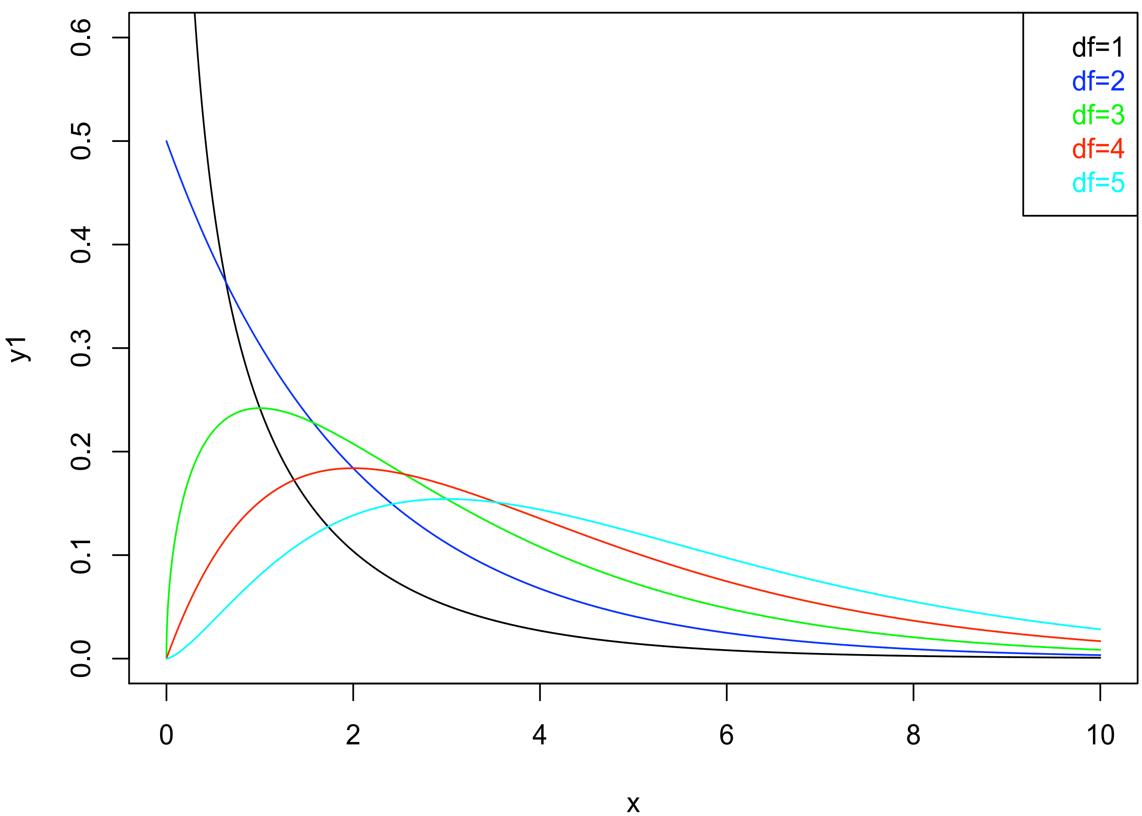
has a singularity at $x = 0$, *and*

$$f_{\chi_2^2}(x) = \frac{1}{2^{2/2}\,\Gamma\left(\dfrac{2}{2}\right)}\, x^{(2/2)-1}\, e^{-x/2} = \frac{1}{2\,\Gamma(1)}\, x^0\, e^{-x/2} = \frac{1}{2}\, e^{-x/2}$$

is simply an exponential decay in $x \geq 0$.

The diagram below shows graphs of the probability density functions $f_{\chi_1^2}, \ldots, f_{\chi_5^2}$. It is generated with `R` with the following command:

```
> x<-seq(0,10,0.001);
> y1<-dchisq(x,df=1); y2<-dchisq(x,df=2); y3<-dchisq(x,df=3); y4<-dchisq(x,df=4); y5<-dchisq(x,df=5);
> plot(x,y1,ylim=c(0,0.6),type="l"); > points(x,y2,type="l",col="blue"); points(x,y3,type="l",col="green");
> points(x,y4,type="l",col="red");points(x,y5,type="l",col="cyan");
> legend("topright",c("df=1","df=2","df=3","df=4","df=5"),text.col=c("black","blue","green","red","cyan"));
```



4 $\mathcal{F}_n^m$ — the distribution of the ratio of two χ^2 random variables

Definition 4.1
*Let $m, n \in \mathbb{N}$. Let $X_m \sim \chi_m^2$ and $X_n \sim \chi_n^2$ be independent χ^2 random variables with the indicated degrees of freedom. For $m, n \in \mathbb{N}$, the **F distribution with m and n degrees of freedom**, denoted by $\mathcal{F}_n^m$, is the probability distribution of the following random variable:*

$$F := \frac{X_m/m}{X_n/n}.$$

Theorem 4.2
The probability density function of the F distribution $\mathcal{F}_n^m$ with m and n degrees of freedom is given by:

$$f_{\mathcal{F}_n^m}(\zeta) \;=\; \left(m^{m/2} \cdot n^{n/2} \cdot \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\,\Gamma\left(\frac{n}{2}\right)} \right) \cdot \frac{\zeta^{(m/2)-1}}{(m\,\zeta + n)^{(m+n)/2}}, \quad \text{for } \zeta \geq 0.$$

Remark 4.3
The "F" in "F distribution" commemorates the renowned statistician Sir Ronald Fisher.
Let $T = \dfrac{Z}{\sqrt{X/n}}$ *be a Student t ratio, i.e.* Z *and* X *are independent random variables with* $Z \sim \mathcal{N}(0,1)$ *and* $X \sim \chi^2_n$.
Then, $Z^2 \sim \chi^2_1$. *Hence,* $T^2 = \dfrac{Z^2}{X/n} = \dfrac{Z^2/1}{X/n} \sim \mathcal{F}^1_n$, *the F distribution with* $m = 1$ *and* n *degrees of freedom. We will derive the probability density function for the distribution of* T *by using that of* T^2 *as given by Theorem 4.2.*

Proof of Theorem 4.2: We first find the probability density function for X_m/X_n. Now,

$$X_m \sim \chi^2_m \implies f_{X_m}(x) = \frac{1}{2^{m/2}\,\Gamma\!\left(\dfrac{m}{2}\right)}\, x^{(m/2)-1}\, e^{-x/2},$$

$$X_n \sim \chi^2_m \implies f_{X_n}(x) = \frac{1}{2^{n/2}\,\Gamma\!\left(\dfrac{n}{2}\right)}\, x^{(n/2)-1}\, e^{-x/2}$$

By Theorem A.1,

$$\begin{aligned}
f_{X_m/X_n}(x) &= \int_0^\infty |\zeta| \cdot f_{X_n}(x) \cdot f_{X_m}(x\,\zeta)\,\mathrm{d}\zeta \\
&= \int_0^\infty \zeta \cdot \left(\frac{1}{2^{n/2}\,\Gamma\!\left(\dfrac{n}{2}\right)}\, \zeta^{(n/2)-1}\, e^{-\zeta/2} \right) \cdot \left(\frac{1}{2^{m/2}\,\Gamma\!\left(\dfrac{m}{2}\right)}\, (x\,\zeta)^{(m/2)-1}\, e^{-(x\,\zeta)/2} \right) \mathrm{d}\zeta \\
&= \frac{1}{2^{(m+n)/2} \cdot \Gamma\!\left(\dfrac{n}{2}\right) \cdot \Gamma\!\left(\dfrac{m}{2}\right)} \cdot \left(x^{(m/2)-1} \right) \cdot \left(\int_0^\infty \zeta^{\frac{m+n}{2}-1} \cdot \exp\!\left(-\,\frac{1+x}{2} \cdot \zeta \right) \mathrm{d}\zeta \right)
\end{aligned}$$

Now, recall again that the probability density function of the $\Gamma(r,\lambda)$ distribution is given by:

$$f_{\Gamma(r,\lambda)}(\zeta) \;=\; \frac{\lambda^r}{\Gamma(r)}\, \zeta^{r-1}\, e^{-\lambda\zeta}, \quad \text{for } \zeta \geq 0.$$

In particular,

$$1 \;=\; \int_0^\infty f_{\Gamma(r,\lambda)}(\zeta)\,\mathrm{d}\zeta \;=\; \int_0^\infty \frac{\lambda^r}{\Gamma(r)}\,\zeta^{r-1}\,e^{-\lambda\zeta}\,\mathrm{d}\zeta, \quad \text{which implies} \quad \int_0^\infty \zeta^{r-1}\,e^{-\lambda\zeta}\,\mathrm{d}\zeta \;=\; \frac{\Gamma(r)}{\lambda^r}.$$

We now see that

$$\begin{aligned}
f_{X_m/X_n}(x) &= \frac{1}{2^{(m+n)/2} \cdot \Gamma\!\left(\frac{n}{2}\right) \cdot \Gamma\!\left(\frac{m}{2}\right)} \cdot \left(x^{(m/2)-1} \right) \cdot \left(\int_0^\infty \zeta^{\frac{m+n}{2}-1} \cdot \exp\!\left(-\,\frac{1+x}{2} \cdot \zeta \right) \mathrm{d}\zeta \right) \\
&= \frac{1}{2^{(m+n)/2} \cdot \Gamma\!\left(\frac{n}{2}\right) \cdot \Gamma\!\left(\frac{m}{2}\right)} \cdot \left(x^{(m/2)-1} \right) \cdot \frac{\Gamma\!\left(\frac{m+n}{2}\right)}{\left(\frac{1+x}{2}\right)^{(m+n)/2}} \\
&= \frac{\Gamma\!\left(\frac{m+n}{2}\right)}{\Gamma\!\left(\frac{m}{2}\right) \cdot \Gamma\!\left(\frac{n}{2}\right)} \cdot \frac{x^{(m/2)-1}}{(1+x)^{(m+n)/2}}
\end{aligned}$$

Lastly, note that for any $\alpha > 0$ and random variable Y, we have:

$$F_{\alpha Y}(y) \;=\; P(\alpha Y \leq y) \;=\; P\!\left(Y \leq \frac{1}{\alpha}\,y \right) \;=\; F_Y\!\left(\frac{1}{\alpha}\,y \right).$$

Hence,

$$f_{\alpha Y}(y) \;=\; \frac{\mathrm{d}}{\mathrm{d}y}\,F_{\alpha\,Y}(y) \;=\; \frac{\mathrm{d}}{\mathrm{d}y}\,F_Y\!\left(\frac{1}{\alpha}\,y\right) \;=\; F'_Y\!\left(\frac{1}{\alpha}\,y\right) \cdot \frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{1}{\alpha}\,y\right) \;=\; \frac{1}{\alpha}\,f_Y\!\left(\frac{1}{\alpha}\,y\right).$$

Consequently,

$$\begin{aligned}
f_{\frac{X_m/m}{X_n/n}}(x) &= f_{\left(\frac{n}{m}\right)X_m/X_n}(x) = \frac{m}{n}\cdot f_{X_m/X_n}\left(\frac{m}{n}\,x\right) \\
&= \frac{m}{n}\cdot\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\cdot\Gamma\left(\frac{n}{2}\right)}\cdot\frac{\left(\frac{m}{n}x\right)^{(m/2)-1}}{\left(1+\frac{m}{n}x\right)^{(m+n)/2}} \\
&= \frac{m}{n}\cdot\left(\frac{m}{n}\right)^{(m/2)-1}\cdot n^{(m+n)/2}\cdot\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\cdot\Gamma\left(\frac{n}{2}\right)}\cdot\frac{x^{(m/2)-1}}{(n+mx)^{(m+n)/2}} \\
&= \left(m^{m/2}\cdot n^{n/2}\cdot\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\cdot\Gamma\left(\frac{n}{2}\right)}\right)\cdot\frac{x^{(m/2)-1}}{(mx+n)^{(m+n)/2}}
\end{aligned}$$

This completes the proof of Theorem 4.2. □

5 The Student t distribution with $n \in \mathbb{N}$ degrees of freedom

Definition 5.1
*The **Student** t **distribution with** $n \in \mathbb{N}$ **degrees of freedom** is the probability distribution of a random variable T_n of the form*

$$T_n = \frac{Z}{\sqrt{\dfrac{X}{n}}},$$

where Z and X are independent random variables, Z is a standard normal random variable, and X is a chi-square random variable with $n \in \mathbb{N}$ degrees of freedom.

Lemma 5.2
The probability density function f_{T_n} of the Student t distribution is an even function, i.e. $f_{T_n}(-t) = f_{T_n}(t)$, for any $t \in \mathbb{R}$.

Proof By definition of the Student t distribution, $T_n = \dfrac{Z}{\sqrt{X/n}}$, where $Z \sim \mathcal{N}(0,1)$ and $X \sim \chi^2_n$ are independent random variables. Consequently, $-T_n = \dfrac{-Z}{\sqrt{X/n}}$ also has the Student t distribution, and thus $f_{(-T_n)}(t) = f_{T_n}(t)$ for all $t \in \mathbb{R}$. On the other hand,

$$F_{(-T_n)}(t) = P(-T_n \leq t) = P(T_n \geq -t) = 1 - P(T_n \leq -t)$$

Differentiating with respect to t yields:

$$\begin{aligned}
f_{(-T_n)}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}F_{(-T_n)}(t) = \frac{\mathrm{d}}{\mathrm{d}t}(1 - P(T_n \leq -t)) = -\frac{\mathrm{d}}{\mathrm{d}t}P(T_n \leq -t) = -\left(\frac{\mathrm{d}}{\mathrm{d}(-t)}P(T_n \leq -t)\right)\cdot\frac{\mathrm{d}(-t)}{\mathrm{d}t} \\
&= -f_{T_n}(-t)\,\frac{\mathrm{d}(-t)}{\mathrm{d}t} = f_{T_n}(-t).
\end{aligned}$$

Thus, we have shown:

$$f_{T_n}(-t) = f_{(-T_n)}(t) = f_{T_n}(t).$$

□

Theorem 5.3
The probability density function of a random variable T_n having the Student t distribution with $n \in \mathbb{N}$ degrees of freedom is given by:

$$f_{T_n}(t) = \left(\frac{1}{\sqrt{n\pi}}\cdot\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\right)\cdot\frac{1}{\left(1+\dfrac{t^2}{n}\right)^{(n+1)/2}}, \quad \text{for} \quad -\infty < t < \infty\,.$$

PROOF Note that $T_n^2 = \dfrac{Z^2}{X/n} = \dfrac{Z^2/1}{X/n} \sim F_n^1$. Hence,

$$f_{T_n^2}(t) \;=\; \frac{n^{n/2}\,\Gamma\!\left(\frac{n+1}{2}\right)}{\Gamma\!\left(\frac{1}{2}\right)\,\Gamma\!\left(\frac{n}{2}\right)}\, t^{-1/2}\,\frac{1}{(t+n)^{(n+1)/2}}\,, \quad \text{for } t>0.$$

Since $f_{T_n}(t)$ is an even function, we have, for $t>0$,

$$\begin{aligned} F_{T_n}(t) &= P(T_n \le t) \;=\; \frac{1}{2} + P(0 \le T_n \le t) \;=\; \frac{1}{2} + \frac{1}{2}P(-t \le T_n \le t) \;=\; \frac{1}{2} + \frac{1}{2}P\big(0 \le T_n^2 \le t\big) \\ &= \frac{1}{2} + \frac{1}{2}F_{T_n^2}\big(t^2\big) \end{aligned}$$

Differentiating with respect to t yields:

$$\begin{aligned} f_{T_n}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}F_{T_n}(t) \;=\; \frac{1}{2}F'_{T_n^2}\big(t^2\big)\,\frac{\mathrm{d}}{\mathrm{d}t}\big(t^2\big) \;=\; t\cdot f_{T_n^2}\big(t^2\big) \;=\; t\cdot\frac{n^{n/2}\,\Gamma\!\left(\frac{n+1}{2}\right)}{\Gamma\!\left(\frac{1}{2}\right)\,\Gamma\!\left(\frac{n}{2}\right)}\cdot\big(t^2\big)^{-1/2}\cdot\frac{1}{(t^2+n)^{(n+1)/2}} \\ &= \frac{n^{n/2}\,\Gamma\!\left(\frac{n+1}{2}\right)}{\Gamma\!\left(\frac{1}{2}\right)\,\Gamma\!\left(\frac{n}{2}\right)}\cdot\frac{1}{n^{(n+1)/2}\left(1+\dfrac{t^2}{n}\right)^{(n+1)/2}} \;=\; \left(\frac{1}{\sqrt{n\,\pi}}\cdot\frac{\Gamma\!\left(\frac{n+1}{2}\right)}{\Gamma\!\left(\frac{n}{2}\right)}\right)\cdot\frac{1}{\left(1+\dfrac{t^2}{n}\right)^{(n+1)/2}} \end{aligned}$$

This completes the proof of Theorem 5.3 □

A The Probability Density Function of the Quotient of Two Random Variables

Theorem A.1
A

B A Technical Result

Theorem B.1
Let $X_1, X_2, \ldots, X_n$ be independent standard normal random variables with common mean μ and (finite) variance $\sigma^2 > 0$. Define:

$$\overline{X} \;:=\; \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{and} \qquad S^2 \;:=\; \frac{1}{n-1}\sum_{i=1}^{n}\big(X_i - \overline{X}\big)^2\,.$$

Then,

- *$\overline{X}$ and S^2 are independent random variables.*
- *$\dfrac{n-1}{\sigma^2}\,S^2$ has a chi-square distribution with $(n-1)$ degrees of freedom.*

PROOF Let $Y_i := \dfrac{X_i - \mu}{\sigma}$, $i = 1, \ldots n$. Then, $Y_1, Y_2, \ldots, Y_n \sim \mathcal{N}(0,1)$, and they are independent random variables. Let $A \in \mathbb{R}^{n\times n}$ be any orientation-preserving orthogonal matrix (hence, $A^T \cdot A = I_{n\times n}$ and $\det(A) = 1$) whose n^{th} row is $\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)$. Let $\mathbf{Y}$ be the $\mathbb{R}^n$-valued random variable defined by $\mathbf{Y} := (Y_1, Y_2, \ldots, Y_n)^T$, and let $\mathbf{Z}$ be the $\mathbb{R}^n$-valued random variable defined by $\mathbf{Z} := A\cdot\mathbf{Y}$. Let $Z_1, Z_2, \ldots, Z_n$ denote the component random variables of $\mathbf{Z}$, i.e. $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)^T$. Note that $Z_n = \frac{1}{\sqrt{n}}Y_1 + \frac{1}{\sqrt{n}}Y_2 + \cdots + \frac{1}{\sqrt{n}}Y_n = \sqrt{n}\,\overline{Y}$, where $\overline{Y} := \frac{1}{n}\sum_{i=1}^{n} Y_i$.

For any measurable $\Omega \subset \mathbb{R}^n$,

$$\begin{aligned}
P(\mathbf{Z} \in \Omega) &= P(A \cdot \mathbf{Y} \in \Omega) \;=\; P\big(\mathbf{Y} \in A^{-1}(\Omega)\big) \\
&= \int_{A^{-1}(\Omega)} f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n)\,\mathrm{d}y_1\cdots\mathrm{d}y_n \\
&= \int_{\Omega} f_{Y_1,\ldots,Y_n}\big(A^{-1}\mathbf{Z}\big)\,\det J(g)\,\mathrm{d}z_1\cdots\mathrm{d}z_n, \quad \text{where } g(\mathbf{Z}) := A^{-1}\cdot\mathbf{Z} \\
&= \int_{\Omega} f_{Y_1,\ldots,Y_n}\big(A^{-1}\mathbf{Z}\big)\cdot 1\cdot\mathrm{d}z_1\cdots\mathrm{d}z_n, \quad \text{since } \det\big(A^{-1}\big)=1 \\
&= \int_{\Omega} \frac{1}{\big(\sqrt{2\pi}\big)^n}\exp\left(-\frac{1}{2}\left\|A^{-1}\mathbf{Z}\right\|^2\right)\mathrm{d}z_1\cdots\mathrm{d}z_n, \quad \text{since } Y_1,\ldots,Y_n\sim\mathcal{N}(0,1) \text{ are independent} \\
&= \int_{\Omega} \frac{1}{\big(\sqrt{2\pi}\big)^n}\exp\left(-\frac{1}{2}\left\|\mathbf{Z}\right\|^2\right)\mathrm{d}z_1\cdots\mathrm{d}z_n, \quad \text{since } A^{-1} \text{ is an orthogonal matrix} \\
&= \int_{\Omega} \prod_{i=1}^{n}\frac{\exp\big(-\frac{1}{2}z_i^2\big)}{\sqrt{2\pi}}\,\mathrm{d}z_1\cdots\mathrm{d}z_n,
\end{aligned}$$

which shows that

$$f_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) \;=\; \prod_{i=1}^{n}\frac{\exp\big(-\frac{1}{2}z_i^2\big)}{\sqrt{2\pi}}.$$

Thus, $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables. Lastly, observe that

$$\sum_{i=1}^{n-1} Z_i^2 + n\,\overline{Y}^2 \;=\; \sum_{i=1}^{n-1} Z_i^2 + (Z_n)^2 \;=\; \|\mathbf{Z}\|^2 \;=\; \left\|A^{-1}\mathbf{Z}\right\|^2 \;=\; \|\mathbf{Y}\|^2 \;=\; \sum_{i=1}^{n} Y_i^2 \;=\; \sum_{i=1}^{n}\big(Y_i-\overline{Y}\big)^2 + n\,\overline{Y}^2,$$

which implies

$$\sum_{i=1}^{n}\big(Y_i-\overline{Y}\big)^2 \;=\; \sum_{i=1}^{n-1} Z_j^2.$$

On the other hand, noting that $\overline{Y} = \dfrac{\overline{X}-\mu}{\sigma}$, we have

$$\begin{aligned}
\frac{n-1}{\sigma^2}\,S^2 &= \frac{n-1}{\sigma^2}\left(\frac{1}{n-1}\sum_{i=1}^{n}\big(X_i-\overline{X}\big)^2\right) \;=\; \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \\
&= \sum_{i=1}^{n}\left(\frac{(X_i-\mu)-(\overline{X}-\mu)}{\sigma}\right)^2 \;=\; \sum_{i=1}^{n}\big(Y_i-\overline{Y}\big)^2 \\
&= \sum_{i=1}^{n-1} Z_j^2.
\end{aligned}$$

This proves that $\dfrac{n-1}{\sigma^2}\,S^2 = \sum_{i=1}^{n-1} Z_j^2$ indeed has a chi-square distribution with $(n-1)$ degrees of freedom, since $Z_1,\ldots,Z_{n-1}$ are independent standard normal random variables. We may also conclude that $\overline{X} = \sigma\,\overline{Y}+\mu = \dfrac{\sigma}{\sqrt{n}}\,Z_n+\mu$ and $\dfrac{n-1}{\sigma^2}\,S^2 = \sum_{i=1}^{n-1} Z_j^2$ are indeed independent random variables, since $Z_1, Z_2, \ldots, Z_n$ are independent. $\square$

C The Fourier transform of a probability measure on $\mathbb{R}$ and the characteristic function of an $\mathbb{R}$-valued random variable

Definition C.1 (Fourier transform of a probability measure on $\mathbb{R}$)

The Fourier transform $\widehat{\mu}$ *of a probability measure* μ *on* $\mathbb{R}$ *is the* $\mathbb{C}$*-valued function* $\widehat{\mu}:\mathbb{R}\longrightarrow\mathbb{C}$ *defined on* $\mathbb{R}$ *via*

$$\widehat{\mu}(\theta) \;:=\; E\{\, e^{\mathbf{i}\theta X} \,\} \;=\; \int_{\mathbb{R}} e^{\mathbf{i}\theta x}\,\mu(\mathrm{d}x), \quad \text{for } \theta\in\mathbb{R}.$$

Definition C.2 (Characteristic function of a random variable)

Let X *be an* $\mathbb{R}$*-valued random variable, and* P_X *its distribution measure on* $\mathbb{R}=\mathrm{codomain}(X)$*. The* characteristic function *of* X *is by definition the Fourier transform of* P_X*. Explicitly, the characteristic function of* X *is the function* $\widehat{P}_X:\mathbb{R}\longrightarrow\mathbb{C}$ *defined by:*

$$\widehat{P}_X(\theta) \;=\; \int_{\mathbb{R}} e^{\mathbf{i}\theta x}\,P_X(\mathrm{d}x), \quad \text{for } \theta\in\mathbb{R}.$$

Theorem C.3 (Theorem 13.1, [2])

The Fourier transform $\widehat{\mu}$ *of any probability measure* μ *on* $\mathbb{R}$ *is a bounded and continuous* $\mathbb{C}$*-valued function on* $\mathbb{R}$*, and* $\widehat{\mu}(0)=1$*.*

Remark C.4

The Fourier transform can thus be regarded as a map from the set[1] *of all probability measures on* $\mathbb{R}$ *into the set of all bounded continuous* $\mathbb{C}$*-valued functions defined on* $\mathbb{R}$*.*

Theorem C.5 (Theorem 13.3, [2])

Let X *be an* $\mathbb{R}$*-valued random variable and* $\alpha,\beta\in\mathbb{R}$*. Then, for any* $\theta\in\mathbb{R}$*,*

$$\widehat{P}_{\alpha X+\beta}(\theta) \;=\; e^{\mathbf{i}\beta\theta}\cdot\widehat{P}_X(\alpha\,\theta).$$

PROOF

$$\widehat{P}_{\alpha X+\beta}(\theta) \;=\; E\left\{e^{\mathbf{i}(\alpha X+\beta)\theta}\right\} \;=\; \int_{\mathbb{R}} e^{\mathbf{i}\beta\theta}\cdot e^{\mathbf{i}(\alpha\theta)x}\,P_X(\mathrm{d}x) \;=\; e^{\mathbf{i}\beta\theta}\cdot\int_{\mathbb{R}} e^{\mathbf{i}(\alpha\theta)x}\,P_X(\mathrm{d}x) \;=\; e^{\mathbf{i}\beta\theta}\cdot\widehat{P}_X(\alpha\,\theta)$$

□

Theorem C.6 (Theorem 15.2, [2])

The characteristic function of the sum of two independent $\mathbb{R}$*-valued random variables is the product of their characteristic functions.*

More precisely, if $X,Y:\Omega\longrightarrow\mathbb{R}$ *are independent* $\mathbb{R}$*-valued random variables with respective characteristic functions* $\widehat{P}_X,\widehat{P}_Y:\mathbb{R}\longrightarrow\mathbb{C}$*, then the characteristic function* $\widehat{P}_Z$ *of the random variable* $Z:=X+Y$ *is given in terms of* $\widehat{P}_X$ *and* $\widehat{P}_Y$ *by:*

$$\widehat{P}_Z(\theta) \;=\; \widehat{P}_X(\theta)\cdot\widehat{P}_Y(\theta), \quad \text{for each } \theta\in\mathbb{R}.$$

Theorem C.7 (Theorem 13.2, [2])

Let X *be an* $\mathbb{R}$*-valued random variable and suppose that* $E\{|X|^m\}<\infty$ *for some non-negative integer* m*. Then the Fourier transform* $\widehat{P}_X$ *of the distribution measure* P_X *has continuous derivatives up to order* m*, and*

$$\frac{\mathrm{d}^m}{\mathrm{d}\theta^m}\widehat{P}_X(\theta) \;=\; \mathbf{i}^m E\{X^m e^{\mathbf{i}\,\theta X}\}$$

Corollary C.8

For an $\mathbb{R}$*-valued random variable* X*,*

$$\begin{aligned} E\{|X|\}<\infty &\implies E\{X\} = -\,\mathbf{i}\,\widehat{P}'_X(0),\\ E\{X^2\}<\infty &\implies E\{X^2\} = -\widehat{P}''_X(0). \end{aligned}$$

[1] Note that the set of probability measures on $\mathbb{R}$ does not form a vector space.

D The Fourier transform of the standard normal distribution measure on $\mathbb{R}$

Recall that probability density function of the standard normal (or standard Gaussian) distribution measure $P_{\mathcal{N}(0,1)}$ is

$$f_{\mathcal{N}(0,1)}(z) := \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \textit{for } z \in \mathbb{R}.$$

The Fourier transform $\widehat{f}_{\mathcal{N}(0,1)}$ of $f_{\mathcal{N}(0,1)}$ is

$$\widehat{f}_{\mathcal{N}(0,1)}(\theta) := E\left(e^{\mathbf{i}\,\theta Z}\right) = \int_{-\infty}^{\infty} e^{\mathbf{i}\,\theta z} f_{\mathcal{N}(0,1)}(z)\,\mathrm{d}z = \int_{-\infty}^{\infty} (\cos(\theta z) + \mathbf{i}\,\sin(\theta z)) \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,\mathrm{d}z = e^{-\theta^2/2}, \quad \textit{for } \theta \in \mathbb{R},$$

The function $\widehat{f}_{\mathcal{N}(0,1)}$ is also, by definition, the Fourier transform $\widehat{P}_{\mathcal{N}(0,1)}$ of the standard Gaussian distribution measure $P_{\mathcal{N}(0,1)}$. In other words, $\widehat{P}_{\mathcal{N}(0,1)} = \widehat{f}_{\mathcal{N}(0,1)}$.

E Injectivity and continuity of the Fourier transform on the space of probability measures on $\mathbb{R}$

Theorem E.1 (Injectivity of the Fourier transform on the space of probability measures on $\mathbb{R}$)
If the Fourier transforms of two probability measures on $\mathbb{R}^d$ are equal (as $\mathbb{C}$-valued functions on $\mathbb{R}^d$), then the two probability measures themselves are equal.

See Theorem 14.1, [2].

Remark E.2
Recall that the Fourier transform can be regarded a map from the set of all probability measures on $\mathbb{R}$ into the set of all bounded continuous $\mathbb{C}$-valued functions defined on $\mathbb{R}$. The above injectivity theorem states that this Fourier transform map is injective.

Theorem E.3 (Levy's Continuity Theorem of the Fourier transform)
Let $\{\,\mu_n\,\}_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathbb{R}^d$, $d \geq 1$, and let $\widehat{\mu_n} : \mathbb{R}^d \longrightarrow \mathbb{C}$ be the Fourier transform of μ_n.

- *If μ_n converges weakly to a measure μ, then $\widehat{\mu_n}$ converges pointwise to $\widehat{\mu}$, i.e. $\widehat{\mu_n}(\theta)$ converges to $\widehat{\mu}(\theta)$, for each $\theta \in \mathbb{R}^d$.*
- *If $\widehat{\mu_u}$ converges pointwise to some function $f : \mathbb{R}^d \longrightarrow \mathbb{C}$, and f is continuous at $\mathbf{0} \in \mathbb{R}^d$, then there exists a probability measure μ on $\mathbb{R}^d$ such that $\widehat{\mu} = f$, and μ_n converges weakly to μ.*

See Theorem 19.1, [2].

References

[1] Ewens, W. J., and Grant, G. R. *Statistical Methods in Bioinformations*, second ed. Statistics for Biology and Health. Springer, New York, 2005.

[2] Jacod, J., and Protter, P. *Probability Essentials.* Springer-Verlag, New York, 2004. Universitext.

[3] Larsen, R. J., and Marx, M. L. *An Introductio to Mathematical Statistics and its Applications*, fourth ed. Pearson/Prentice Hall, Upper Saddle River, New Jersey, 2006.

[4] Wasserman, L. A. *All of Statistics: A Concise Course in Statistical Inference*, first ed. Spinger Texts in Statistics. Springer-Verlag, New York, 2004.