Probability Theory

Wilhelm Stannat

Technische Universität Darmstadt

Winter Term 2007/08

Third part - corrected version

This text is a summary of the lecture on Probability Theory held at the TU Darmstadt in Winter Term 2007/08.

Please email all misprints and mistakes to
stannat@mathematik.tu-darmstadt.de

Bibliography

1. Bauer, H., *Probability theory*, de Gruyter, 1996.
2. Bauer, H., *Maß- und Integrationstheorie*, de Gruyter, 1996.
3. Billingsley, P., *Probability and Measure*, Wiley, 1995.
4. Billingsley, P., *Convergence of probability measures*, Wiley, 1999.
5. Dudley, R.M., *Real analysis and probability*, Cambridge University Press, 2002.
6. Elstrodt, J., *Maß- und Integrationstheorie*, Springer, 2005.
7. Feller, W., *An introduction to probability theory and its applications*, Vol. 1 & 2, Wiley, 1950.
8. Halmos, P.R., *Measure Theory*, Springer, 1974.
9. Klenke, A., *Wahrscheinlichkeitstheorie*, Springer, 2006.
10. Shiryaev, A.N., *Probability*, Springer, 1996.

2 Independence

4 Joint distribution and convolution

Let $X_i \in \mathcal{L}^1$ i.i.d. Kolmogorov's law of large numbers implies that

$$\frac{1}{n} \underbrace{\sum_{i=1}^{n} X_i(\omega)}_{=:S_n} \xrightarrow{n\to\infty} \mathbb{E}[X_1] \quad P\text{-a.s.}$$

hence

$$\int f(x)\, d\left(P \circ \left(\frac{S_n}{n}\right)^{-1}\right)(x) = \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right]$$

$$\xrightarrow[n\to\infty]{\text{(Lebesgue)}} f\big(\mathbb{E}[X_1]\big) = \int f(x)\, d\delta_{\mathbb{E}[X_1]}(x) \qquad \forall f \in C_b(\mathbb{R})$$

i.e., the distribution of $\frac{S_n}{n}$ converges weakly to $\delta_{\mathbb{E}[X_1]}$. This is not surprising, because at least for $X_i \in \mathcal{L}^2$

$$\operatorname{var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \sum_{i=1}^{n} \underbrace{\operatorname{var}(X_i)}_{=\operatorname{var}(X_1)} \xrightarrow{n\to\infty} 0.$$

We will see later that if we rescale S_n appropriately, namely $\frac{1}{\sqrt{n}} S_n$, so that $\operatorname{var}\big(\frac{1}{\sqrt{n}} S_n\big) = \operatorname{var}(X_1)\big)$, the sequence of distributions of $\frac{1}{\sqrt{n}} S_n$ is asymptotically distributed as a normal distribution.
One problem in this context is: How to calculate the distribution of S_n? Since S_n is a function of $X_1, \dots, X_n$, we need to consider their *joint* distribution in the sense of the following definition:

Definition 4.1. Let $X_1, \dots, X_n$ be real-valued r.v. on $(\Omega, \mathcal{A}, P)$. Then the distribution $\bar{\mu} := P \circ \bar{X}^{-1}$ of the transformation

$$\bar{X} : \Omega \to \mathbb{R}^n, \quad \omega \mapsto \big(X_1(\omega), \dots, X_n(\omega)\big)$$

under P is said to be the *joint distribution* of $X_1, \dots, X_n$.
Note that $\bar{\mu}$ is a probability measure on $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big)$ with $\bar{\mu}(\bar{A}) = P[\bar{X} \in \bar{A}]$ for all $\bar{A} \in \mathcal{B}(\mathbb{R}^n)$.

Remark 4.2. *(i) $\bar{\mu}$ is well-defined, because $\bar{X} : \Omega \to \mathbb{R}^n$ is $\mathcal{A}/\mathcal{B}(\mathbb{R}^n)$-measurable.*

Proof:

$$\begin{aligned}\mathcal{B}(\mathbb{R}^n) &= \sigma\big(\{A_1 \times \cdots \times A_n \mid A_i \in \mathcal{B}(\mathbb{R})\}\big) \\ &= \sigma\big(\{A_1 \times \cdots \times A_n \mid A_i =]-\infty, x_i]\,,\ x_i \in \mathbb{R}\}\big)\end{aligned}$$

and

$$\bar{X}^{-1}(A_1 \times \cdots \times A_n) = \bigcap_{i=1}^{n} \underbrace{\{X_i \in A_i\}}_{\in \mathcal{A}} \quad \in \mathcal{A}$$

which implies the measurability of the transformation $\bar{X}$ (see Remark 1.3.2 (ii))

(ii) *Proposition 1.11.5 implies that $\bar{\mu}$ is uniquely determined by*

$$\bar{\mu}(A_1 \times \cdots \times A_n) = P\Big(\bigcap_{i=1}^{n} \{X_i \in A_i\}\Big).$$

Example 4.3. (i) Let X, Y be r.v., uniformly distributed on $[0,1]$. Then

- X, Y independent $\Rightarrow$ joint distribution = uniform distribution on $[0,1]^2$
- $X = Y$ $\Rightarrow$ joint distribution = uniform distribution on the diagonal

(ii) Let X, Y be independent, $N(m, \sigma^2)$ distributed. The following Proposition shows that the joint distribution of X and Y has the density

$$f(x,y) = \frac{1}{2\pi\sigma^2} \cdot \exp\Big[-\frac{1}{2\sigma^2} \cdot \big((x-m)^2 + (y-m)^2\big)\Big]$$

which is a particular example of a *2-dimensional normal distribution*.

In the case $m = 0$ it follows that

$$R := \sqrt{X^2 + Y^2},$$

$$\Phi := \arctan \frac{Y}{X},$$

are independent and

Φ has a uniform distribution on $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$,

R has a density $\begin{cases} \frac{r}{\sigma^2}\exp\big(-\frac{r^2}{2\sigma^2}\big) & \text{if } r \geqslant 0 \\ 0 & \text{if } r < 0. \end{cases}$

Definition 4.4. (Products of probability spaces) The product of measurable spaces $(\Omega_i, \mathcal{A}_i)$, $i = 1, \ldots n$, is defined as the measurable space

$$\Omega := \Omega_1 \times \ldots \times \Omega_n$$

endowed with the smallest σ-algebra

$$\mathcal{A} := \sigma\{A_1 \times \ldots \times A_n \mid A_i \in \mathcal{A}_i, 1 \le i \le n\}$$

generated by measurable cylindrical sets. $\mathcal{A}$ is said to be the *product σ-algebra* of $\mathcal{A}_i$ (notation: $\bigotimes_{i=1}^n \mathcal{A}_i$).

Let P_i, $i = 1, \ldots, n$, be probability measures on $(\Omega_i, \mathcal{A}_i)$. Then there exists a unique probability measure P on the product space $(\Omega, \mathcal{A})$ satisfying

$$P(A_1 \times \cdots \times A_n) = P_1(A_1) \cdot \ldots \cdot P_n(A_n)$$

for every measurable cylindrical set. P is called the *product measure* of P_i (notation: $\bigotimes_{i=1}^n P_i$).

(Uniqueness of P follows from 1.11.5, existence later!)

Proposition 4.5. *Let $X_1, \ldots, X_n$ be r.v. on $(\Omega, \mathcal{A}, P)$ with distributions $\mu_1, \ldots, \mu_n$ and joint distribution $\bar{\mu}$. Then*

$$X_1, \ldots, X_n \text{ independent} \quad \Leftrightarrow \quad \bar{\mu} = \bigotimes_{i=1}^n \mu_i,$$

(i.e., $\bar{\mu}(A_1 \times \cdots \times A_n) = \prod_i \mu_i(A_i)$ if $A_i \in \mathcal{B}(\mathbb{R})$).

In this case:

(i) $\bar{\mu}$ is uniquely determined by $\mu_1, \ldots, \mu_n$.

(ii)

$$\int \varphi(x_1, \ldots, x_n) \, \mathrm{d}\bar{\mu}(x_1, \ldots, x_n) = \int \left(\cdots \left(\int \varphi(x_1, \ldots, x_n) \, \mu_{i_1}(\mathrm{d}x_{i_1}) \right) \cdots \right) \mu_{i_n}(\mathrm{d}x_{i_n}).$$

for all $\mathcal{B}(\mathbb{R}^n)$-measurable functions $\varphi : \mathbb{R}^n \to \bar{R}$ with $\varphi \ge 0$ or φ $\bar{\mu}$-integrable.

(iii) If μ_i is absolutely continuous with density f_i, $i = 1, \ldots, n$, then $\bar{\mu}$ is absolutely continuous with density

$$\bar{f}(\bar{x}) := \prod_{i=1}^n f_i(x_i).$$

Proof. The equivalence is obvious.

(i) Obvious from part (ii) of the previous Remark 4.2.

(ii) See text books on measure theory.

(iii) $\bar{f}$ is nonnegative and measurable on $\mathbb{R}^n$ w.r.t. $\mathcal{B}(\mathbb{R}^n)$, and

$$\int_{\mathbb{R}^n} \bar{f}(\bar{x})\,\mathrm{d}\bar{x} = \prod_{i=1}^{n} \int_{\mathbb{R}} f_i(x_i)\,\mathrm{d}x_i = 1.$$

Hence,

$$\check{\mu}(A) := \int_A \bar{f}(\bar{x})\,\mathrm{d}\bar{x}, \quad A \in \mathcal{B}(\mathbb{R}^n),$$

defines a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. For $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ it follows that

$$\begin{aligned}\bar{\mu}(A_1 \times \cdots \times A_n) &= \prod_{i=1}^{n} \mu_i(A_i) = \prod_{i=1}^{n} \int_{A_i} f_i(x_i)\,\mathrm{d}x_i \\ &\overset{\text{(ii)}}{=} \int 1_{A_1 \times \cdots \times A_n}(\bar{x}) \cdot \bar{f}(\bar{x})\,\mathrm{d}\bar{x} = \check{\mu}(A_1 \times \cdots \times A_n).\end{aligned}$$

Hence $\bar{\mu} = \check{\mu}$ by 1.11.5. □

Let $X_1, \dots, X_n$ be independent, $S_n := X_1 + \cdots + X_n$
How to calculate the distribution of S_n with the help of the distribution of X_i?
In the following denote by $T_x : \mathbb{R}^1 \to \mathbb{R}^1$, $y \mapsto x + y$, the translation by $x \in \mathbb{R}$.

Proposition 4.6. *Let X_1, X_2 be independent r.v. with distributions μ_1, μ_2. Then:*

(i) The distribution of $X_1 + X_2$ is given by the convolution

$$\begin{aligned}\mu_1 * \mu_2 &:= \int \mu_1(\mathrm{d}x_1)\, \mu_2 \circ T_{x_1}^{-1}, \textit{i.e.} \\ \mu_1 * \mu_2(A) &== \int 1_A(x_1 + x_2)\, \mu_1(\mathrm{d}x_1)\, \mu_2(\mathrm{d}x_2) \\ &= \int \mu_1(\mathrm{d}x_1)\, \mu_2(A - x_1) \qquad \forall\, A \in \mathcal{B}(\mathbb{R}^1).\end{aligned}$$

(ii) If one of the distributions μ_1, μ_2 is absolutely continuous, e.g. μ_2 with density f_2, then $\mu_1 * \mu_2$ is absolutely continuous again with density

$$\begin{aligned}f(x) &:= \int \mu_1(\mathrm{d}x_1)\, f_2(x - x_1) \\ &\Big(= \int f_1(x_1) \cdot f_2(x - x_1)\,\mathrm{d}x_1 =: (f_1 * f_2)(x) \quad \textit{if } \mu_1 = f_1\,\mathrm{d}x_1. \Big)\end{aligned}$$

Proof. (i) Let $A \in \mathcal{B}(\mathbb{R})$, and define $\bar{A} := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 + x_2 \in A\})$. Then

$$\begin{aligned} P[X_1 + X_2 \in A] &= P\big[(X_1, X_2) \in \bar{A}\big] = (\mu_1 \otimes \mu_2)(\bar{A}) \\ &= \iint 1_{\bar{A}}(x_1, x_2)\, \mathrm{d}(\mu_1 \otimes \mu_2)(x_1, x_2) \\ &= \iint 1_A(x_1 + x_2)\, \mathrm{d}(\mu_1 \otimes \mu_2)(x_1, x_2) \\ &= \int \left(\int 1_{A - x_1}(x_2)\, \mu_2(\mathrm{d}x_2) \right) \mu_1(\mathrm{d}x_1) \\ &= \int \mu_2(A - x_1)\, \mu_1(\mathrm{d}x_1) = (\mu_1 * \mu_2)(A). \end{aligned}$$

(ii)

$$\begin{aligned} (\mu_1 * \mu_2)(A) &= \int \mu_1(\mathrm{d}x_1)\, \mu_2(A - x_1) = \int \mu_1(\mathrm{d}x_1) \int_{A - x_1} f_2(x_2)\, \mathrm{d}x_2 \\ &\overset{\substack{\text{change of variable} \\ x - x_1 = x_2}}{=} \int \mu_1(\mathrm{d}x_1) \int_A f_2(x - x_1)\, \mathrm{d}x \\ &\overset{4.5}{=} \int_A \left(\int \mu_1(\mathrm{d}x_1)\, f_2(x - x_1) \right) \mathrm{d}x. \end{aligned}$$

□

Example 4.7.

(i) Let X_1, X_2 be independent r.v. with Poisson-distribution π_{λ_1} and π_{λ_2}. Then $X_1 + X_2$ has Poisson-distribution $\pi_{\lambda_1 + \lambda_2}$, because

$$\begin{aligned} (\pi_{\lambda_1} * \pi_{\lambda_2})(n) &= \sum_{k=0}^{n} \pi_{\lambda_1}(k) \cdot \pi_{\lambda_2}(n - k) = e^{-(\lambda_1 + \lambda_2)} \sum_{k=0}^{n} \frac{\lambda_1^k}{k!} \cdot \frac{\lambda_2^{n-k}}{(n-k)!} \\ &= e^{-(\lambda_1 + \lambda_2)} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} \cdot \lambda_1^k \lambda_2^{n-k} = e^{-(\lambda_1 + \lambda_2)} \cdot \frac{(\lambda_1 + \lambda_2)^n}{n!}. \end{aligned}$$

(ii) Let X_1, X_2 be independent r.v. with normal distributions $N(m_i, \sigma_i^2)$, $i = 1, 2$. Then $X_1 + X_2$ has normal distribution $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$, because $f_{m_1 + m_2, \sigma_1^2 + \sigma_2^2} = f_{m_1, \sigma_1^2} * f_{m_2, \sigma_2^2}$ (Exercise!)

(iii) The *Gamma distribution* $\Gamma_{\alpha,p}$ is defined through its density $\gamma_{\alpha,p}$ given by

$$\gamma_{\alpha,p}(x) = \begin{cases} \frac{1}{\Gamma(p)} \cdot \alpha^p x^{p-1} e^{-\alpha x} & \text{if } x > 0 \\ 0 & \text{if } x \leqslant 0 \end{cases}$$

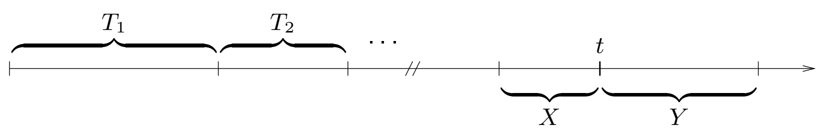
If X_1, X_2 are independent with distribution Γ_{α, p_i}, $i = 1, 2$, then $X_1 + X_2$ has distribution $\Gamma_{\alpha, p_1 + p_2}$. (Exercise!)

In the particular case $p_i = 1$: The sum $S_n = T_1 + \ldots + T_n$ of independent r.v. T_i with exponential distribution with parameter α has Gamma-distribution $\Gamma_{n,\alpha}$, i.e.

$$\gamma_{\alpha,n}(x) = \begin{cases} \frac{\alpha^n}{(n-1)!} \cdot e^{-\alpha x} x^{n-1} & \text{if } x > 0 \\ 0 & \text{if } x \leqslant 0. \end{cases}$$

Example 4.8 (The waiting time paradox)**.** Let $T_1, T_2, \ldots$ be independent, exponentially distributed waiting times with parameter $\alpha > 0$, so that in particular

$$\mathbb{E}[T_i] = \int_0^\infty x \cdot \alpha e^{-\alpha x} \,\mathrm{d}x = \cdots = \frac{1}{\alpha}.$$



Question: Fix some time t. How long on average is the remaining waiting time until the next event, i.e., how big is $\mathbb{E}[Y]$?
Answer: $\mathbb{E}[Y] = \frac{1}{\alpha}$, and

$$\mathbb{E}[X + Y] = \frac{1}{\alpha}(1 - e^{-\alpha t}) \approx \frac{1}{\alpha} \text{ for large } t\,.$$

More precisely:

(i) X, Y are independent.

(ii) Y has exponential distribution with parameter α.

(iii) X has exponential distribution with parameter α, "conditioned on" $[0, t]$, i.e.:

$$P[X \geqslant s] = e^{-\alpha s} \quad \forall\, 0 \leqslant s \leqslant t,$$
$$P[X = t] = e^{-\alpha t};$$

In particular,

$$\mathbb{E}[X] = \int_0^t s \cdot \alpha e^{-\alpha s} \,\mathrm{d}s + t \cdot e^{-\alpha t} = \cdots = \frac{1}{\alpha}(1 - e^{-\alpha t}).$$

Proof. Let us first determine the joint distribution of X and Y: Fix $0 \leqslant x \leqslant t$ and

$y \geqslant 0$. Then for $S_n := T_1 + \cdots + T_n$, $S_0 := 0$:

$$
\begin{aligned}
&P[X \geqslant x,\ Y \geqslant y] \\
&\quad = P\Big(\dot{\bigcup_{n\in\mathbb{N}}} \{S_n + x \leqslant t,\ S_{n+1} - y \geqslant t\}\Big) \\
&\quad = P[T_1 \geqslant y + t] + \sum_{n=1}^{\infty} P\big[S_n \leqslant t - x,\ T_{n+1} \geqslant y + t - S_n\big] \\
&\quad = e^{-\alpha(t+y)} + \sum_{n=1}^{\infty} \iint_{\substack{\{(s,r)\,|\,s\leqslant t-x,\\ r\geqslant y+t-s\}}} \gamma_{\alpha,n}(s)\cdot \alpha e^{-\alpha r}\,\mathrm{d}s\,\mathrm{d}r \\
&\quad = e^{-\alpha(t+y)} + \sum_{n=1}^{\infty} \int_{\{s\,|\,s\leqslant t-x\}} \gamma_{\alpha,n}(s)\cdot e^{-\alpha(y+t-s)}\,\mathrm{d}s \\
&\quad = e^{-\alpha(t+y)}\bigg(1 + \int_0^{t-x} e^{\alpha s} \underbrace{\sum_{n=1}^{\infty} \gamma_{\alpha,n}(s)}_{\overset{(*)}{=}\alpha}\ \mathrm{d}s\bigg) \\
&\quad = e^{-\alpha(t+y)}\bigg(1 + \int_0^{t-x} \alpha e^{\alpha s}\,\mathrm{d}s\bigg) \\
&\quad = e^{-\alpha(t+y)}\cdot e^{\alpha(t-x)} = e^{-\alpha y}\cdot e^{-\alpha x}.
\end{aligned}
$$

Consequently:

(i) If $x = 0$: Y ist exponentially distributed with parameter α.

(ii) If $y = 0$: X ist exponentially distributed with parameter α, conditioned on $[0, t]$.

(iii) X, Y are independent.

We have used in line six that:

$$
\sum_{n=1}^{\infty} \gamma_{\alpha,n}(s) = \sum_{n=1}^{\infty} \frac{\alpha^n}{(n-1)!}\cdot s^{n-1} e^{-\alpha s} = \alpha e^{-\alpha s} \sum_{n=1}^{\infty} \frac{(\alpha s)^{n-1}}{(n-1)!} = \alpha e^{-\alpha s} e^{\alpha s} = \alpha. \ \Box
$$

5 Characteristic functions

Let $\mathcal{M}_+^1(\mathbb{R}^n)$ be the set of all probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

For given $\mu \in \mathcal{M}_+^1(\mathbb{R}^n)$ define its *characteristic function* as the complex-valued function $\hat{\mu} : \mathbb{R}^n \to \mathbb{C}$ defined by

$$
\hat{\mu}(u) := \int e^{i\langle u,y\rangle}\,\mu(dy) := \int \cos(\langle u,y\rangle)\,\mu(dy) + i\int \sin(\langle u,y\rangle)\,\mu(dy)\,.
$$

Proposition 5.1. *Let $\mu \in \mathcal{M}_+^1(\mathbb{R}^n)$. Then* *(i)* $\hat{\mu}(0) = 1$.

(ii) $|\hat{\mu}| \leqslant 1$.

(iii) $\hat{\mu}(-u) = \overline{\hat{\mu}(u)}$.

(iv) $\hat{\mu}$ *is uniformly continuous.*

(v) $\hat{\mu}$ *is* positive definite, *i.e. for all* $c_1, \ldots, c_m \in \mathbb{C}$, $u_1, \ldots, u_m \in \mathbb{R}^n$, $m \geqslant 1$*:*

$$\sum_{j,k=1}^{m} c_j \bar{c}_k \cdot \hat{\mu}(u_j - u_k) \geqslant 0.$$

Proof. Exercise. □

Proposition 5.2 (Uniqueness theorem)**.** *Let* $\mu_1, \mu_2 \in \mathcal{M}_+^1(\mathbb{R}^n)$ *with* $\hat{\mu}_1 = \hat{\mu}_2$*. Then* $\mu_1 = \mu_2$*.*

Proposition 5.3 (Bochner's theorem)**.** *Let* $\varphi : \mathbb{R}^n \to \mathbb{C}$ *be a continuous, positive definite function with* $\varphi(0) = 1$*. Then there exists one (and only one)* $\mu \in \mathcal{M}_+^1(\mathbb{R}^n)$ *with* $\hat{\mu} = \varphi$*.*

Proposition 5.4 (Lévy's continuity theorem)**.** *Let* $(\mu_m)_{m \in \mathbb{N}}$ *be a sequence in* $\mathcal{M}_+^1(\mathbb{R}^n)$*. Then*

(i) $\lim_{m\to\infty} \mu_m = \mu$ *weakly implies* $\lim_{m\to\infty} \hat{\mu}_m = \hat{\mu}$ *uniformly on every compact subset of* $\mathbb{R}^n$*.*

(ii) Conversely, if $(\hat{\mu}_m)_{m\in\mathbb{N}}$ *converges pointwise to some function* $\varphi : \mathbb{R}^n \to \mathbb{C}$ *which is continuous in* $u = 0$*, then there exists a unique* $\mu \in \mathcal{M}_+^1(\mathbb{R}^n)$ *such that* $\hat{\mu} = \varphi$ *and* $\lim_{m\to\infty} \mu_m = \mu$ *weakly.*

Proof. See Satz 15.23 in Klenke. □

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $X : \Omega \to \mathbb{R}^n$ be $\mathcal{A}/\mathcal{B}(\mathbb{R}^n)$-measurable. Let P_X $(:= P \circ X^{-1})$ be the distribution of X. Then

$$\varphi_X(u) := \hat{P}_X(u) = \int e^{i\langle u,y\rangle}\, P_X(dy) = \int e^{i\langle u,X\rangle} dP = E\left[e^{i\langle u,X\rangle}\right]$$

is said to be the *characteristic function of* X.

Remark 5.5. $X_1, \ldots, X_n$ *are independent if and only if*

$$\underbrace{\hat{P}_{(X_1,\ldots,X_n)}}_{=\varphi_{(X_1,\ldots,X_n)}}(u_1, \ldots, u_n) = \prod_{j=1}^{n} \underbrace{\hat{P}_{X_j}(u_j)}_{=\varphi_{X_j}(u_j)} \quad \big(= (P_{X_1} \otimes \hat{\cdots} \otimes P_{X_n})(u_1, \ldots, u_n)\big),$$

i.e.: $\hat{P}_{(X_1,\ldots,X_n)} = \prod_{j=1}^{n} \hat{P}_{X_j}$.

Proposition 5.6. *Let $X_1, \ldots, X_n$ be independent r.v., $\alpha \in \mathbb{R}$ and $S := \alpha \sum_{k=1}^n X_k$. Then for all $u \in \mathbb{R}$:*

$$\varphi_S(u) = \prod_{k=1}^n \varphi_{X_k}(\alpha u).$$

Proof.

$$\varphi_S(u) = \int e^{iuS} \, \mathrm{d}P = \int \prod_{k=1}^n e^{i\alpha u X_k} \, \mathrm{d}P \overset{\text{Indep.}}{=} \prod_{k=1}^n \int e^{i\alpha u X_k} \, \mathrm{d}P = \prod_{k=1}^n \varphi_{X_k}(\alpha u) \square$$

Proposition 5.7. *For all $u \in \mathbb{R}^n$:*

$$\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int e^{i\langle u,y\rangle} e^{-\frac{1}{2}|y|^2} \, \mathrm{d}y = e^{-\frac{1}{2}\|u\|^2}.$$

Proof. See Satz 15.12 in Klenke. ☐

Example 5.8. (i) $\hat{\delta}_a(u) = e^{iua}$.

(ii) Let $\mu := \sum_{i=1}^\infty \alpha_i \delta_{a_i}$ ($\alpha_i \geq 0$, $\sum_{i=1}^\infty \alpha_i = 1$). Then

$$\hat{\mu}(u) = \sum_{i=1}^\infty \alpha_i e^{iua_i}, \quad u \in \mathbb{R}^n.$$

Examples:

a) **Binomial distribution** $\beta_n^p = \sum_{k=1}^n \binom{n}{k} p^k q^{n-k} \, \delta_k$ Then for all $u \in \mathbb{R}$:

$$\hat{\beta}_n^p(u) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \cdot e^{iuk} = (q + pe^{iu})^n.$$

b) **Poisson distribution** $\pi_\alpha = \sum_{n=0}^\infty e^{-\alpha} \, \frac{\alpha^n}{n!} \, \delta_n$. Then for all $u \in \mathbb{R}$:

$$\hat{\pi}_\alpha(u) = e^{-\alpha} \sum_{n=0}^\infty \underbrace{\frac{\alpha^n}{n!} \cdot e^{iun}}_{=\frac{(\alpha e^{iu})^n}{n!}} = e^{\alpha(e^{iu}-1)}.$$

6 Central limit theorem

Definition 6.1. Let $X_1, X_2, \ldots \in \mathcal{L}^2$ be independent r.v., $S_n := X_1 + \cdots + X_n$ and

$$S_n^* := \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{var}(S_n)}} \qquad \text{("standardization")}$$

(so that in particular $\mathbb{E}[S_n^*] = 0$ and $\mathrm{var}(S_n^* = 1)$). The sequence $X_1, X_2, \ldots$ of r.v. is said to have the *central limit property* (CLP), if

$$\lim_{n\to\infty} P_{S_n^*} = N(0,1) \quad \text{weakly,}$$

or equivalently (by the Portmanteau theorem)

$$\lim_{n\to\infty} P[S_n^* \leqslant b] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} e^{-\frac{x^2}{2}}\, \mathrm{d}x = \Phi(b) \quad \forall\, b \in \mathbb{R}.$$

Proposition 6.2. *(Central limit theorem) Let* $X_1, X_2, \ldots \in \mathcal{L}^2$ *be independent r.v.,* $\sigma_n^2 := \mathrm{var}(X_n) > 0$ *and*

$$s_n := \Big(\sum_{k=1}^{n} \sigma_k^2\Big)^{\frac{1}{2}}.$$

Assume that $(X_n)_{n\in\mathbb{N}}$ *satisfies* Lindeberg's condition

$$\lim_{n\to\infty} \underbrace{\sum_{k=1}^{n} \int_{\left\{\frac{|X_k - \mathbb{E}[X_k]|}{s_n} \geqslant \varepsilon\right\}} \Big(\frac{X_k - \mathbb{E}[X_k]}{s_n}\Big)^2 \,\mathrm{d}P}_{=:L_n(\varepsilon)} = 0 \qquad \forall\, \varepsilon > 0\,. \tag{L}$$

Then $(X_n)_{n\in\mathbb{N}}$ *has the CLP.*

Remark 6.3. *(i)* $(X_n)_{n\in\mathbb{N}}$ *i.i.d.* $\Rightarrow$ $(X_n)_{n\in\mathbb{N}}$ *satisfies* (L)*.*

Proof: *Let* $m := \mathbb{E}[X_n]$*,* $\sigma^2 := \mathrm{var}(X_n)$*. Then* $s_n^2 = n\sigma^2$*, so that*

$$L_n(\varepsilon) = \sigma^{-2} \int_{\{|X_1 - m| \geqslant \varepsilon\sqrt{n}\sigma\}} (X_1 - m)^2 \,\mathrm{d}P \xrightarrow[\text{Lebesgue}]{n\to\infty} 0.$$

(ii) The following stronger condition, known as Lyapunov's condition*, is often easier to check in applications:*

$$\exists\, \delta > 0 : \lim_{n\to\infty} \frac{\displaystyle\sum_{k=1}^{n} \mathbb{E}\Big[\big|X_k - \mathbb{E}[X_k]\big|^{2+\delta}\Big]}{s_n^{2+\delta}} = 0\,. \tag{Lya}$$

To see that Lyapunov's condition implies Lindeberg's condition note that for all $\varepsilon > 0$*:*

$$\begin{aligned} &\big|X_k - \mathbb{E}[X_k]\big| &&\geqslant \varepsilon s_n \\ \Rightarrow\quad &\big|X_k - \mathbb{E}[X_k]\big|^{2+\delta} &&\geqslant \big|X_k - \mathbb{E}[X_k]\big|^2 \cdot (\varepsilon s_n)^\delta \end{aligned}$$

and therefore

$$L_n(\varepsilon) \leqslant \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\Big[\big|X_k - \mathbb{E}[X_k]\big|^{2+\delta}\Big].$$

(iii) Let (X_n) be bounded and suppose that $s_n \to \infty$. Then (X_n) satisfies Lyapunov's condition for any $\delta > 0$, because

$$\begin{aligned} & |X_k| \leqslant \frac{\alpha}{2} \\ \Rightarrow \quad & \big|X_k - \mathbb{E}[X_k]\big| \leqslant \alpha \\ \Rightarrow \quad & \frac{1}{s_n^{2+\delta}} \sum_{k=1}^n \mathbb{E}\Big[\big|X_k - \mathbb{E}[X_k]\big|^{2+\delta}\Big] \\ & \quad \leqslant \frac{\alpha^\delta}{s_n^{2+\delta}} \underbrace{\sum_{k=1}^n \mathbb{E}\Big[\big|X_k - \mathbb{E}[X_k]\big|^2\Big]}_{=s_n^2} = \left(\frac{\alpha}{s_n}\right)^\delta . \end{aligned}$$

Lemma 6.4. *Suppose that (X_n) satisfies Lindeberg's condition. Then*

$$\lim_{n\to\infty} \max_{1\leqslant k\leqslant n} \frac{\sigma_k}{s_n} = 0 . \tag{2.1}$$

Proof. For all $1 \leqslant k \leqslant n$

$$\begin{aligned} \left(\frac{\sigma_k}{s_n}\right)^2 &= \int \left(\frac{X_k - \mathbb{E}[X_k]}{s_n}\right)^2 \mathrm{d}P \leqslant \int_{\left\{\frac{|X_k - \mathbb{E}[X_k]|}{s_n} \geqslant \varepsilon\right\}} \left(\frac{X_k - \mathbb{E}[X_k]}{s_n}\right)^2 \mathrm{d}P + \varepsilon^2 \\ &\leqslant L_n(\varepsilon) + \varepsilon^2 . \end{aligned}$$

□

The proof of Proposition 6.2 requires some further preparations.

Lemma 6.5. *For all $t \in \mathbb{R}$ and $n \in \mathbb{N}$:*

$$\left| e^{it} - 1 - \frac{it}{1!} - \frac{(it)^2}{2!} - \cdots - \frac{(it)^{n-1}}{(n-1)!} \right| \leqslant \frac{|t|^n}{n!} .$$

Proof. Define $f(t) := e^{it}$, then $f^{(k)}(t) = i^k e^{it}$, and the Taylor series expansion, applied to real and imaginary part, implies that

$$\left| e^{it} - 1 - \cdots - \frac{(it)^{n-1}}{(n-1)!} \right| = \big|R_n(t)\big|$$

with

$$\big|R_n(t)\big| = \left| \frac{1}{(n-1)!} \int_0^t (t-s)^{n-1} i^n e^{is} \, \mathrm{d}s \right| \leqslant \frac{1}{(n-1)!} \int_0^{|t|} s^{n-1} \, \mathrm{d}s = \frac{|t|^n}{n!} .$$

□

Proposition 6.6. *Let $X \in \mathcal{L}^2$. Then $\varphi_X(u) = \int e^{iuX} \, dP$ is two times continuously differentiable with*

$$\varphi_X'(u) = i \int X e^{iuX} \, \mathrm{d}P , \qquad \varphi_X''(u) = -\int X^2 e^{iuX} \, \mathrm{d}P .$$

In particular

$$\varphi_X'(0) = i \cdot \mathbb{E}[X], \quad \varphi_X''(0) = -\mathbb{E}[X^2], \quad |\varphi_X''| \leqslant \mathbb{E}[X^2].$$

Moreover, for all $u \in \mathbb{R}$

$$\varphi_X(u) = 1 + iu \cdot \mathbb{E}[X] + \frac{1}{2} \cdot \theta(u) u^2 \cdot \mathbb{E}[X^2]$$

with $\big|\theta(u)\big| \leqslant 1$ *and* $\theta(u) \in \mathbb{C}$.

Proof. Clearly,

$$(e^{iuX})' = iX \cdot e^{iuX}, \quad (e^{iuX})'' = -X^2 e^{iuX}, \quad |e^{iuX}| = 1\,.$$

Now, Lebesgue's dominated convergence theorem implies all assertions up to the last one. For the proof of the last assertion note that the previous lemma implies in the case $n = 2$ that

$$|e^{iuX} - 1 - iuX| \leqslant \frac{1}{2} \cdot u^2 X^2.$$

Integration w.r.t. P now implies that

$$\left|\int e^{iuX} - 1 - iuX \,\mathrm{d}P\right| = \big|\varphi_X(u) - 1 - iu \cdot \mathbb{E}[X]\big| \leqslant \frac{1}{2} \cdot u^2 \cdot \mathbb{E}[X^2]. \qquad \square$$

From now on assume that $X_1, X_2, \dots \in \mathcal{L}^2$ are independent and

$$\mathbb{E}[X_n] = 0 \;\forall\, n, \quad \sigma_n^2 := \operatorname{var}(X_n) > 0, \quad s_n = \Big(\sum_{k=1}^n \sigma_k^2\Big)^{\frac{1}{2}}.$$

Proposition 6.7. *Suppose that*

$$(a) \qquad \lim_{n\to\infty} \max_{1\leq k\leq n} \frac{\sigma_k}{s_n} = 0 \qquad \textit{and}$$

$$(b) \qquad \lim_{n\to\infty} \sum_{k=1}^n \left(\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right) = -\frac{1}{2}u^2 \qquad \forall\, u \in \mathbb{R}\,.$$

Then (X_n) *has the CLP.*

Proof. It is sufficient to show that

$$\lim_{n\to\infty} \prod_{k=1}^n \varphi_{X_k}\left(\frac{u}{s_n}\right) = e^{-\frac{1}{2}u^2}. \tag{2.2}$$

because for $S_n^* = \frac{1}{s_n}\sum_{k=1}^n X_k$ we have that

$$\varphi_{S_n^*}(u) = \prod_{k=1}^n \varphi_{X_k}\left(\frac{u}{s_n}\right),$$

and $\varphi_{S_n^*}(u) \xrightarrow{n\to\infty} e^{-\frac{1}{2}u^2} = \widehat{N(0,1)}(u)$ pointwise, implies by Lévy's continuity theorem that $\lim_{n\to\infty} P_{S_n^*} = N(0,1)$ weakly.

For the proof of (2.6) we need to show that for all $u \in \mathbb{R}$

$$\lim_{n\to\infty}\left(\prod_{k=1}^n \varphi_{X_k}\left(\frac{u}{s_n}\right) - \underbrace{\prod_{k=1}^n \exp\left[\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right]}_{=\exp[\sum\cdots]\to\exp[-\frac{1}{2}u^2]}\right) = 0\,.$$

To this end fix $u \in \mathbb{R}$ and note that $|\varphi_{X_k}| \leqslant 1$, hence

$$\left|\exp\left[\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right]\right| = \exp\left[\mathrm{Re}\,\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right] \leqslant 1.$$

Note that for $a_1, \ldots, a_n, b_1, \ldots, b_n \in \{z \in \mathbb{C} \mid |z| \leqslant 1\}$

$$\begin{aligned}\left|\prod_{k=1}^n a_k - \prod_{k=1}^n b_k\right| &= \big|(a_1 - b_1)\cdot a_2\cdots a_n + b_1\cdot(a_2 - b_2)\cdot a_3\cdots a_n + \ldots \\ &\qquad + b_1\cdots b_{n-1}\cdot(a_n - b_n)\big| \\ &\leqslant \sum_{k=1}^n |a_k - b_k|.\end{aligned}$$

Consequently,

$$\begin{aligned}&\left|\prod_{k=1}^n \varphi_{X_k}\left(\frac{u}{s_n}\right) - \prod_{k=1}^n \exp\left[\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right]\right| \\ &\quad\leqslant \sum_{k=1}^n \left|\varphi_{X_k}\left(\frac{u}{s_n}\right) - \exp\left[\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1\right]\right| =: D_n.\end{aligned}$$

If we define $z_k := \varphi_{X_k}(\frac{u}{s_n}) - 1$, we can write

$$D_n = \sum_{k=1}^n |z_k + 1 - e^{z_k}|.$$

Note that $\mathbb{E}[X_k] = 0$ and $\mathbb{E}[X_k^2] = \sigma_k^2$. The previous proposition now implies that for all k

$$|z_k| \leqslant \frac{1}{2}\left(\frac{u}{s_n}\right)^2 \sigma_k^2.$$

For $\varepsilon > 0$ we can find $\delta > 0$ such that

$$|z + 1 - e^z| \leqslant \varepsilon|z| \qquad \forall\, |z| < \delta\,.$$

Let $n_0 \in \mathbb{N}$ be such that for all $n \geqslant n_0$

$$\frac{u^2}{2}\left(\frac{\sigma_k}{s_n}\right)^2 < \delta$$

for all $n \geq n_0$. Then, for all $n \geqslant n_0$

$$D_n \leqslant \varepsilon \sum_{k=1}^{n} |z_k| \leqslant \varepsilon \frac{u^2}{2} \sum_{k=1}^{n} \frac{\sigma_k^2}{s_n^2} = \varepsilon \cdot \frac{u^2}{2}.$$

Consequently, $\lim_{n\to\infty} D_n = 0$. □

Proof of Proposition 6.2. It remains to show that Lindeberg's condition implies

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left(\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1 \right) = -\frac{1}{2} \cdot u^2.$$

W.l.o.g. assume that $\mathbb{E}[X_n] = 0$ for all $n \in \mathbb{N}$. Let $u \in \mathbb{R}$, $n \in \mathbb{N}$, $1 \leqslant k \leqslant n$. Lemma 6.5 implies that

$$Y_k := \left| \exp\left(i \cdot \frac{u}{s_n} \cdot X_k \right) - 1 - \underbrace{i \cdot \frac{u}{s_n} \cdot X_k}_{\mathbb{E}[\ldots]=0} + \frac{1}{2} \cdot \frac{u^2}{s_n^2} \cdot X_k^2 \right| \leqslant \frac{1}{6} \left| \frac{u}{s_n} \cdot X_k \right|^3,$$

and

$$Y_k \leqslant \left| \exp\left(i \cdot \frac{u}{s_n} \cdot X_k \right) - 1 - i \cdot \frac{u}{s_n} \cdot X_k \right| + \frac{1}{2} \cdot \frac{u^2}{s_n^2} \cdot X_k^2 \leqslant \frac{u^2}{s_n^2} \cdot X_k^2.$$

With these notations

$$\begin{aligned} \left| \sum_{k=1}^{n} \left(\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1 \right) + \frac{1}{2} \cdot u^2 \right| &= \left| \sum_{k=1}^{n} \left(\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1 + \frac{1}{2} \cdot \frac{u^2}{s_n^2} \cdot \sigma_k^2 \right) \right| \\ &\leqslant \sum_{k=1}^{n} \mathbb{E}[Y_k]. \end{aligned}$$

For $\delta > 0$

$$\begin{aligned} \mathbb{E}[Y_k] &= \int_{\{|X_k| \geqslant \delta s_n\}} Y_k \, \mathrm{d}P + \int_{\{|X_k| < \delta s_n\}} Y_k \, \mathrm{d}P \\ &\leqslant \frac{u^2}{s_n^2} \int_{\{|X_k| \geqslant \delta s_n\}} X_k^2 \, \mathrm{d}P + \frac{|u|^3}{6 s_n^3} \int_{\{|X_k| < \delta s_n\}} |X_k|^3 \, \mathrm{d}P. \end{aligned}$$

Note that

$$\frac{1}{s_n^3} \int_{\{|X_k| < \delta s_n\}} |X_k|^3 \, \mathrm{d}P \leqslant \frac{\delta}{s_n^2} \int X_k^2 \, \mathrm{d}P = \delta \cdot \frac{\sigma_k^2}{s_n^2},$$

so that for $\varepsilon > 0$ and $\delta > 0$ with $\frac{|u|^3}{6}\delta < \frac{\varepsilon}{2}$, we obtain that

$$\sum_{k=1}^{n} \mathbb{E}[Y_k] \leqslant u^2 \sum_{k=1}^{n} \int_{\{|\frac{X_k}{s_n}|\geqslant\delta\}} \left(\frac{X_k}{s_n}\right)^2 \mathrm{d}P + \frac{|u|^3}{6} \cdot \delta \underbrace{\sum_{k=1}^{n} \frac{\sigma_k^2}{s_n^2}}_{=1}$$

$$\leqslant u^2 L_n(\delta) + \frac{\varepsilon}{2}\,.$$

Note that $u^2 L_n(\delta) < \frac{\varepsilon}{2}$ for large n, so that

$$\lim_{n\to\infty} \sum_{k=1}^{n} E[Y_k] = 0\,,$$

and thus

$$\lim_{n\to\infty} \left(\sum_{k=1}^{n} \left[\varphi_{X_k}\left(\frac{u}{s_n}\right) - 1 \right] + \frac{1}{2} \cdot u^2 \right) = 0.$$

Now the assertion follows from Proposition 6.7. □

Example 6.8 (Applications). (i) **"Ruin probability"**

Consider a portfolio of n contracts of a risc insurance (e.g. car insurance, fire insurance, health insurance, ...). Let $X_i \geqslant 0$ be the claim size (or claim severity) of the ith contract, $1 \leqslant i \leqslant n$. We assume that $X_1, \ldots, X_n \in \mathcal{L}^2$ are i.i.d. with $m := \mathbb{E}[X_i]$ and $\sigma^2 := \mathrm{var}(X_i)$.

Suppose the insurance holder has to pay he following premium

$$\Pi := m + \lambda\sigma^2$$
$$= \text{average claim size} + \text{safety loading}.$$

After some fixed amount of time:

Income: $n\Pi$

Expenditures: $S_n = \sum_{i=1}^{n} X_i$.

Suppose that K is the initial capital of the insurance company. What is the probability $P(R)$, where

$R := \{S_n > K + n\Pi\}$ denotes the ruin ?

We assume here that:

- No interest rate.
- Payments due only at the end of the time period.

Let

$$S_n^* := \frac{S_n - nm}{\sqrt{n}\sigma}.$$

The central limit theorem implies for large n that $S_n^* \sim N(0,1)$, so that

$$\begin{aligned} P(R) &= P\left[S_n^* > \frac{K + n\Pi - nm}{\sqrt{n}\sigma}\right] = P\left[S_n^* > \frac{K + n\lambda\sigma^2}{\sqrt{n}\sigma}\right] \\ &\overset{\text{CLT}}{\approx} 1 - \Phi\Bigg(\underbrace{\frac{K + n\lambda\sigma^2}{\sqrt{n}\sigma}}_{\xrightarrow{n\to\infty}\infty}\Bigg), \end{aligned}$$

where Φ denotes the distribution of the standard normal distribution. Note that the ruin probability decreases with an increasing number of contracts.

Example
Assume that $n = 2000$, $\sigma = 60$, $\lambda = 0.5‰$.

(a) $K = 0 \Rightarrow P(R) \approx 1 - \Phi(1.342) \approx 9\%$.

(b) $K = 1500 \Rightarrow P(R) \approx 3\%$.

How large do we have to choose n in order to let the probability of ruin $P(R)$ fall below $1‰$?
Answer: $\Phi(\dots) \geqslant 0.999$, hence $n \geqslant 10\,611$.

(ii) **Stirling's formula**
Remark: Stirling proved the following formula

$$n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} \tag{2.3}$$

in the year 1730 and De Moivre used it in his proof of the CLT for Bernoulli experiments.

Conversely, in 1977, Weng provided an independent proof of the formula, using the CLT (note that we did not use Stirling's formula in our proof of the CLT). Here is Weng's proof:

Let $X_1, X_2, \dots$ be i.i.d. with distribution π_1, i.e.,

$$P_{X_n} = e^{-1} \sum_{k=0}^{\infty} \frac{1}{k!} \delta_k\,.$$

Then $S_n := X_1 + \dots + X_n$ has Poisson distribution π_n, i.e.,

$$P_{S_n} = e^{-n} \sum_{k=0}^{\infty} \frac{n^k}{k!} \delta_k,$$

and in particular $\mathbb{E}[S_n] = \mathrm{var}(S_n) = n$. As usual, define

$$S_n^* := \frac{S_n - n}{\sqrt{n}},$$

so that $S_n^* = t_n \circ S_n$ for $t_n(x) := \frac{x-n}{\sqrt{n}}$. Then

$$\int f \, \mathrm{d}P_{S_n^*} = \mathbb{E}\big[f(S_n^*)\big] = \mathbb{E}\big[(f \circ t_n)(S_n)\big] = \int f \circ t_n \, \mathrm{d}\underbrace{P_{S_n}}_{=\pi_n}$$

In particular, for

$$f_\infty(x) := x^- \quad \big(= (-x) \vee 0\big)$$

it follows that

$$\begin{aligned}
\int f_\infty \, \mathrm{d}P_{S_n^*} &= \int_{\mathbb{R}} \underbrace{f_\infty\left(\frac{x-n}{\sqrt{n}}\right)}_{\begin{cases} = 0 \text{ if } x > n \\ = \frac{n-x}{\sqrt{n}} \text{ if } x \leqslant n \end{cases}} \pi_n(\mathrm{d}x) = e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} \cdot \underbrace{\frac{n-k}{\sqrt{n}}}_{=f_\infty(k)} \\
&= \frac{e^{-n}}{\sqrt{n}} \cdot \left(n + \sum_{k=1}^{n} \frac{n^k (n-k)}{k!}\right) \\
&= \frac{e^{-n}}{\sqrt{n}} \cdot \Bigg(n + \underbrace{\sum_{k=1}^{n} \frac{n^{k+1}}{k!} - \frac{n^k}{(k-1)!}}_{= \frac{n^{n+1}}{n!} - \frac{n^1}{0!}}\Bigg) = \frac{e^{-n} \cdot n^{n+\frac{1}{2}}}{n!}.
\end{aligned}$$

Moreover,

$$\int f_\infty \, \mathrm{d}N(0,1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} (-x) \cdot e^{-\frac{x^2}{2}} \, \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \Big|_{-\infty}^{0} = \frac{1}{\sqrt{2\pi}}.$$

Hence, Stirling's formula (2.7) would follow, once we have shown that

$$\int f_\infty \, \mathrm{d}P_{S_n^*} \xrightarrow{n \to \infty} \int f_\infty \, \mathrm{d}N(0,1). \tag{2.4}$$

Note that this is not implied by the weak convergence in the CLT since f_∞ is continuous but unbounded. Hence, we consider for given $m \in \mathbb{N}$

$$f_m := f_\infty \wedge m \quad \big(\in C_b(\mathbb{R})\big).$$

The CLT now implies that

$$\int f_m \, \mathrm{d}P_{S_n^*} \xrightarrow{n \to \infty} \int f_m \, \mathrm{d}N(0,1).$$

Define $g_m := f_\infty - f_m \quad (\geq 0)$. (2.8) then follows from a "3ε-argument", once we have shown that

$$
\begin{aligned}
&(0 \leqslant) \quad \int g_m \,\mathrm{d}P_{S_n^*} \leqslant \frac{1}{m} \qquad \forall\, m, \\
&(0 \leqslant) \quad \int g_m \,\mathrm{d}N(0,1) \leqslant \frac{1}{m} \qquad \forall m.
\end{aligned}
$$

The first inequality follows from

$$
\begin{aligned}
\int g_m \,\mathrm{d}P_{S_n^*} \quad &= \quad \int_{]-\infty,-m[} \big(|x| - m\big) \,\mathrm{d}P_{S_n^*} \leqslant \int_{]-\infty,-m]} |x| \,\mathrm{d}P_{S_n^*} \\
&\overset{\frac{|x|}{m} \geqslant 1}{\leqslant} \frac{1}{m} \int_{]-\infty,-m]} x^2 \,\mathrm{d}P_{S_n^*} \leqslant \frac{1}{m} \cdot \underbrace{\mathrm{var}(S_n^*)}_{=1},
\end{aligned}
$$

the second inequality can be shown similarly.