10.1 $X_1, X_2, \dots, X_n$ unabhängig mit $X_i \sim \mathrm{Uni}([0,\Pi])$

(a) $$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{(Stichprobenmittel)}$$

$$\mathbb{E}[\overline{X}] = \mathbb{E}[X_1] = \frac{\Pi}{2}$$

$T_1 := c_1 \overline{X}$ ist erwartungstreu für Π, falls $\mathbb{E}[T_1] = \Pi$

$$\mathbb{E}[T_1] \underset{\substack{\uparrow \\ T_1 := c_1 \overline{X} \\ \text{\& Linearität}}}{=} c_1 \, \mathbb{E}[\overline{X}] = c_2 \cdot \frac{\Pi}{2} \overset{!}{=} \Pi \implies c_2 = 2$$

$$\implies T_1 = 2\overline{X}$$

(b) ML-Funktion: $L(\Pi; \underbrace{x_1, \dots, x_n}_{\vec{x}}) := \prod_{i=1}^{n} f(x_i)$

mit $f(x) = I_{[0,\Pi]}(x) \cdot \frac{1}{\Pi}$

ML-Schätzer: $\delta(\underbrace{x_1, \dots, x_n}_{\vec{x}})$ mit $L(\delta(\vec{x}); \vec{x})$ maximal

- $L(\Pi; \vec{x}) = 0$, falls es ein x_i mit $x_i > \Pi$ gibt,
- Es gilt also $\forall i: x_i \leq \Pi$, d.h. $\max(x_1, \dots, x_n) \leq \Pi$

↳ damit: $L(\Pi; \vec{x}) = \frac{1}{\Pi^n}$

↳ $\frac{1}{\Pi^n}$ streng monoton fallend für $\Pi \to \infty$

↳ maximal für $\Pi = \max(x_1, \dots, x_n)$

↳ ML-Schätzer $\delta = \max(X_1, \dots, X_n)$

c) $T_2 := c_2 S$ soll erwartungstreu für Π sein, d.h. $\mathbb{E}[T_2] = c_2 \mathbb{E}[S] \overset{!}{=} \Pi$

$\hookrightarrow$ Dichte von S: $f_S(t) = n\, F(t)^{n-1} \cdot f(t)$ (siehe HA 10.2)

$$= n \cdot \frac{t^{n-1}}{\Pi^n} \cdot I_{[0,\Pi]}(t)$$

(alternativ explizit nachrechnen)

$$\hookrightarrow \mathbb{E}[S] = \int_0^{\Pi} t \cdot n \cdot \frac{t^{n-1}}{\Pi^n}\, dt = \frac{n}{n+1} \cdot \frac{\Pi^{n+1}}{\Pi^n} = \frac{n}{n+1}\Pi$$

[Der ML-Schätzer ist also nicht erwartungstreu]

$$\hookrightarrow c_2 = \frac{n+1}{n}$$

$$\hookrightarrow T_2 = \frac{n+1}{n} \max(X_1, \dots, X_n)$$

korrigiert ~~größte~~ gesehene Zahl somit leicht nach oben.

(d) Für T Schätzer für Parameter θ

$$mse_\theta(T) := \mathbb{E}_\theta[(T-\theta)^2]$$

$$= \mathbb{E}_\theta[(T - \mathbb{E}[T] + \mathbb{E}_\theta[T] - \theta)^2]$$

$b := \mathbb{E}[T] - \theta$ ist konstant; wird als Bias bezeichnet

$$= \mathbb{E}_\theta[(T-\mathbb{E}_\theta[T])^2 + 2(T-\mathbb{E}_\theta[T])\cdot b + b^2]$$

$$= \mathbb{E}_\theta[(T-\mathbb{E}_\theta[T])^2] + 2\cdot b\cdot \underbrace{\mathbb{E}_\theta[T-\mathbb{E}_\theta[T]]}_{} + b^2$$

$$\overset{\text{Linearität}}{=} \mathbb{E}_\theta[T] - \mathbb{E}_\theta[\underbrace{\mathbb{E}_\theta[T]}_{\text{konstant}}] = 0$$

θ bestimmt (implizit) die Verteilung des Schätzers; um sich daran zu erinnern, wird das manchmal als Index für Pr, $\mathbb{E}$, usw mitgeschleppt

$$= Var_\theta[T] + b^2$$

$$= Var_\theta[T] + (\mathbb{E}[T] - \theta)^2$$

$\Rightarrow$ für erwartungstreue Schätzer T Bias $= 0$ & damit $mse_\theta(T) = Var_\theta[T]$.

$\circ \rightsquigarrow \quad mse_\Theta(T_1) = Var_\Theta[T_1] \underset{\substack{\uparrow \\ T_1 = 2\overline{X}}}{=} 4\,Var_\Theta[\overline{X}] = \frac{4}{n} Var[X_1]$

$$= \frac{4}{n} \cdot \frac{\Theta^2}{12} = \underline{\underline{\frac{\Theta^2}{3n}}}$$

$\circ \rightsquigarrow \quad mse_\Theta(T_2) = Var_\Theta[T_2] \underset{\uparrow}{=} \left(\frac{n+1}{n}\right)^2 \left(\mathbb{E}[S^2] - \mathbb{E}[S]^2\right)$

$$T_2 = \frac{n+1}{n} \underbrace{\max(X_1, \dots, X_n)}_{=S}$$

$\cdot\ \mathbb{E}[S] = \frac{n}{n+1} \cdot \Theta \quad$ (siehe ©)

$\cdot\ \mathbb{E}[S^2] = \int_0^\Theta t^2 \cdot n \cdot \frac{t^{n-1}}{\Theta^n}\, dt = \frac{n}{n+2}\Theta^2$

$\Rightarrow \quad Var[S] = \frac{n}{n+2}\Theta^2 - \left(\frac{n}{n+1}\right)^2 \Theta^2 = \frac{n}{(n+2)(n+1)^2}\Theta^2$

$\Rightarrow \quad mse_\Theta(T_2) = \frac{1}{n(n+2)}\Theta^2$

$\Rightarrow$ Beide Schätzer sind konsistent im quadr. Mittel, nur T_2 konvergiert schneller

(e) Symmetrisches Konfidenzintervall zum Signifikanzniveau α gesucht,

d.h. δ in Abhängigkeit von α und Stichprobenumfang n gesucht, mit:

$$\Pr[\,|T - \Pi| > \delta\,] \leq \alpha$$

oder äquivalent: $\Pr[\,|T - \Pi| \leq \delta\,] \geq 1 - \alpha$

(vgl. auch HA 9.2 (a-ii))

Für T_1:

$$\Pr[\,|T_1 - \Pi| \leq \delta\,] = \Pr[\,\Pi - \delta \leq T_1 \leq \Pi + \delta\,]$$

$$= \Pr[\,\Pi - \delta - E[T_1] \leq T_1 - E[T_1] \leq \Pi + \delta - E[T_1]\,]$$

$E[T_1] = \Pi$

$$= \Pr[\,-\delta \leq T_1 - E[T_1] \leq \delta\,]$$

$T_1 = 2\overline{X}$

$$= \Pr\left[\,-\frac{\delta}{2} \leq \overline{X} - E[\overline{X}] \leq \frac{\delta}{2}\,\right]$$

$Var[\overline{X}] = \frac{\Pi^2}{12n}$

$$= \Pr\left[\,-\frac{\delta\cdot\sqrt{3n}}{\Pi} \leq \frac{\overline{X} - E[\overline{X}]}{\sqrt{Var[\overline{X}]}} \leq \frac{\delta\cdot\sqrt{3n}}{\Pi}\,\right]$$

$$\approx 2\Phi\left(\delta\cdot\frac{\sqrt{3n}}{\Pi}\right) - 1$$

$$\Rightarrow \quad \frac{\delta \cdot \sqrt{3n}}{\Pi} \approx z_{1-\frac{\alpha}{2}} \quad \Rightarrow \quad \delta \approx \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{3n}} \cdot \Pi$$

• Für T_2: $\Pr[\,|T_2 - \Pi| \le \delta\,] = \Pr[\,\Pi - \delta \le T_2 \le \Pi + \delta\,]$

$$= \Pr[\,T_2 \le \Pi + \delta\,] - \Pr[\,T_2 \underset{\substack{\uparrow \\ T_2 \text{ stetig verteilt}}}{\le} \Pi - \delta\,] = (*)$$

Verteilungsfunktion von T_2: $\Pr[T_2 \le t] = \Pr[\max(X_1, \ldots, X_n) \le \frac{n}{n+1} \cdot t]$

$$= F_\delta\left(\frac{n}{n+1} \cdot t\right) = \begin{cases} 0 & \text{für } t \le 0 \\ 1 & \text{für } t > \frac{n+1}{n} \cdot \Pi \\ \frac{\left(\frac{n}{n+1} \cdot t\right)^n}{\Pi^n} & \text{für } t \in [0, \frac{n+1}{n} \Pi] \end{cases}$$

$$\Rightarrow \Pr[\,|T_2 - \Pi| \le \delta\,] = F_\delta\left(\frac{n}{n+1}(\Pi + \delta)\right) - F_\delta\left(\frac{n}{n+1}(\Pi - \delta)\right)$$

$$\underset{\text{hinreichend}}{\ge} F_\delta\left(\frac{n}{n+1}\Pi\right) - F_\delta\left(\frac{n}{n+1}(\Pi - \delta)\right) \ge 1 - \alpha$$

$$\rightsquigarrow \quad F_\delta\left(\frac{n}{n+1} M\right) - F_\delta\left(\frac{n}{n+1}(M-\delta)\right)$$

$$\underset{\text{OE } (\delta < M)}{=} \quad \frac{\left(\frac{n}{n+1} M\right)^n}{M^n} - \frac{\left(\frac{n}{n+1}(M-\delta)\right)^n}{M^n}$$

$$= \left(\frac{n}{n+1}\right)^n - \left(\frac{n}{n+1}\right)^n \frac{(M-\delta)^n}{M^n} \overset{!}{\geq} 1-\alpha$$

$$\rightsquigarrow \quad \delta \geq \left(1 - \sqrt[n]{1 - (1-\alpha)\left(\frac{n+1}{n}\right)^n}\right) \cdot M$$

Zum Vergleich:

$$1-\sqrt[n]{\alpha} = 1 - e^{\frac{1}{n}\log\alpha} \qquad (\alpha\in[0,1] \rightsquigarrow \log\alpha \leq 0)$$

$$= 1 - \sum_{k=0}^{\infty} \frac{(\frac{1}{n}\cdot\log\alpha)^k}{k}$$

verhält sich asymptotisch wie $\frac{1}{n}$,

$\Rightarrow$ Der Schätzer T_2 ist auch bzgl. dem Konfidenz-intervall besser als T_1.

(f) Test zum Signifikanzniveau α für H_0: $\Pi \geq \Pi_0$ vs. H_1: $\Pi < \Pi_0$

$\rightsquigarrow$ Signifikanzniveau limitiert hier die W'keit,

dass man sich im Fall, dass H_0 gilt, fälschlicherweise doch für H_1 entscheidet.

<u>1. Frage:</u> Wann soll man sich für H_0 bzw. H_1 anhand des Werts von T_i
T_i $(i \in \{1,2\})$ entscheiden?

⇝ Sowohl $T_1 = 2\overline{X}$ als auch $T_2 = \frac{n+1}{n} \max(X_1, \ldots, X_n)$
sind erwartungstreu: $\mathbb{E}[T_i] = \Omega$

⇝ Unter H_0 sollte man größere Werte von T_i erwarten
als unter H_1

⇝ Lehne H_1 ab, falls T_i zu klein

⇝ <u>D.h.</u> Ablehnungskriterium für H_0: $T_i < J_i$ $(i \in \{1,2\})$

„Schwellwert"
(Threshold....)

$\rightsquigarrow$ Analogerechnung wie in ⓔ:

$$\Pr[T_1 < \mathcal{J}_1] = \Pr\left[\frac{T_1 - E[T_1]}{\sqrt{Var[T_1]}} < \frac{\mathcal{J}_1 - \Pi}{\sqrt{\frac{\Pi^2}{3n}}}\right]$$

$T_1 = 2\overline{X}$
ist Summe von ZVn
$\rightsquigarrow$ ZGWS anwendbar

$$\approx \Phi\left(\frac{\mathcal{J}_1 - \Pi}{\Pi}\sqrt{3n}\right)$$

Fehlerw'keit 1. Art: $H_0: \Pi \geq \Pi_0$ ist wahr, nur zufällig tritt "$T_1 < \mathcal{J}_1$" ein $\rightsquigarrow$ diese W'keit soll maximal α betragen:

D.h.: $\forall \underbrace{\Pi \geq \Pi_0}_{H_0}: \Phi\left(\frac{\mathcal{J}_1 - \Pi}{\Pi}\sqrt{3n}\right) \overset{!}{\leq} \alpha$

$\Longleftrightarrow \forall \Pi \geq \Pi_0 \quad \mathcal{J}_1 \leq \Pi + z_\alpha \cdot \Pi \cdot \frac{1}{\sqrt{3n}}$

$\Longleftrightarrow \boxed{\mathcal{J}_1 \leq \Pi_0 + z_\alpha \cdot \Pi_0 \cdot \frac{1}{\sqrt{3n}}}$

Fehlerw'keit 2. Art:

H_1 gilt, es tritt aber zufällig " $T_1 \geq \delta_1$ " ein.

$\leadsto$ D.h. wollen für alle $\mu < \mu_0$ (H_1) die W'keit

$$\Pr_\mu[T_1 \geq \delta_1]$$

beschränken.

$$\Pr_\mu[T_1 \geq \delta_1] = 1 - \Pr_\mu[T_1 < \delta_1]$$

$$= 1 - \Pr_\mu\left[\frac{T_1 - E[T_1]}{\sqrt{Var[T_1]}} < \frac{\delta_1 - \mu}{\sqrt{\mu^2/3n}}\right]$$

$$= 1 - \Phi\left(\frac{\mu_0}{\mu} \cdot (\sqrt{3n} + z_\alpha) - \sqrt{3n}\right)$$

Kleinste obere Schranke für $\Pr_\mu[T_1 \geq \delta_1]$ falls $H_1: \mu < \mu_0$ gilt:

$$\sup_{\mu < \mu_0} (1 - \Phi(\dots)) = 1 - \alpha$$

Für $T_2 = \frac{n+1}{n} \max(X_1, X_2, \dots, X_n)$:

Fehlerw'keit 1. Art:

$$\forall \Pi \geq \Pi_0: \quad P_{\Pi}[T_2 < \sigma_2] \underset{A}{=} \left(\frac{\frac{n}{n+1}\sigma_2}{\Pi}\right)^n \overset{!}{\leq} \alpha$$

$$\sigma_2 \leq \frac{n+1}{n} \Pi$$

sonst sinnlos

$$\Leftrightarrow \forall \Pi \geq \Pi_0: \quad \sigma_2 \leq \frac{n+1}{n} \cdot \Pi \cdot \sqrt[n]{\alpha}$$

$$\Leftrightarrow \quad \sigma_2 \leq \frac{n+1}{n} \Pi_0 \sqrt[n]{\alpha}$$

Fehlerw'keit 2. Art:

$$\sup_{\Pi < \Pi_0} P_{\Pi}[T_2 \geq \sigma_2] = \sup_{\Pi < \Pi_0} \left(1 - \left(\frac{\Pi_0}{\Pi}\right)^n \alpha\right)$$

$$= 1 - \alpha$$

10.2

(a) $U \sim \mathrm{Uni}([0,1])$

$X \sim \mathrm{Uni}([a,b])$

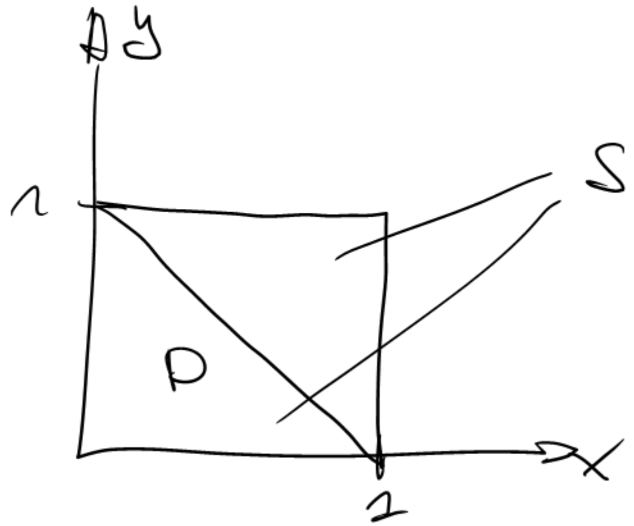
$\rightsquigarrow F_X(t) = \dfrac{t-a}{b-a}$ für $t \in [a,b]$

$\rightsquigarrow F_X^{-1}(t) = (b-a)t + a$

$\rightsquigarrow F_X^{-1}(U) = (b-a)U + a$ ist uniform auf $[a,b]$ verteilt

(was recht offensichtlich ist...)

(b)



(i) $\Pr[\vec{V} \notin D] = \dfrac{vol(D)}{vol(S)} = \dfrac{1}{2}$

$\rightsquigarrow$ Bei n Versuchen $\vec{V}^{(1)}, \vec{V}^{(2)}, \ldots, \vec{V}^{(n)}$

$\rightsquigarrow I_{[\vec{V}^{(i)} \notin D]} \sim Ber(\frac{1}{2})$

$\rightsquigarrow \mathbb{E}[\sum I_{[\vec{V}^{(i)} \notin D]}] = \frac{n}{2}$

verworfene Punkte

maximal umständliche Begründung

(ii) Sei $A \subseteq D$ eine Borel-Menge $(A \in \mathcal{B}(D))$

$$\Pr[\vec{V} \in A \mid \vec{V} \in D] = \frac{\Pr[\vec{V} \in A]}{\Pr[\vec{V} \in D]} = \frac{\frac{\mathrm{vol}(A)}{\mathrm{vol}(S)}}{\frac{\mathrm{vol}(D)}{\mathrm{vol}(S)}}$$

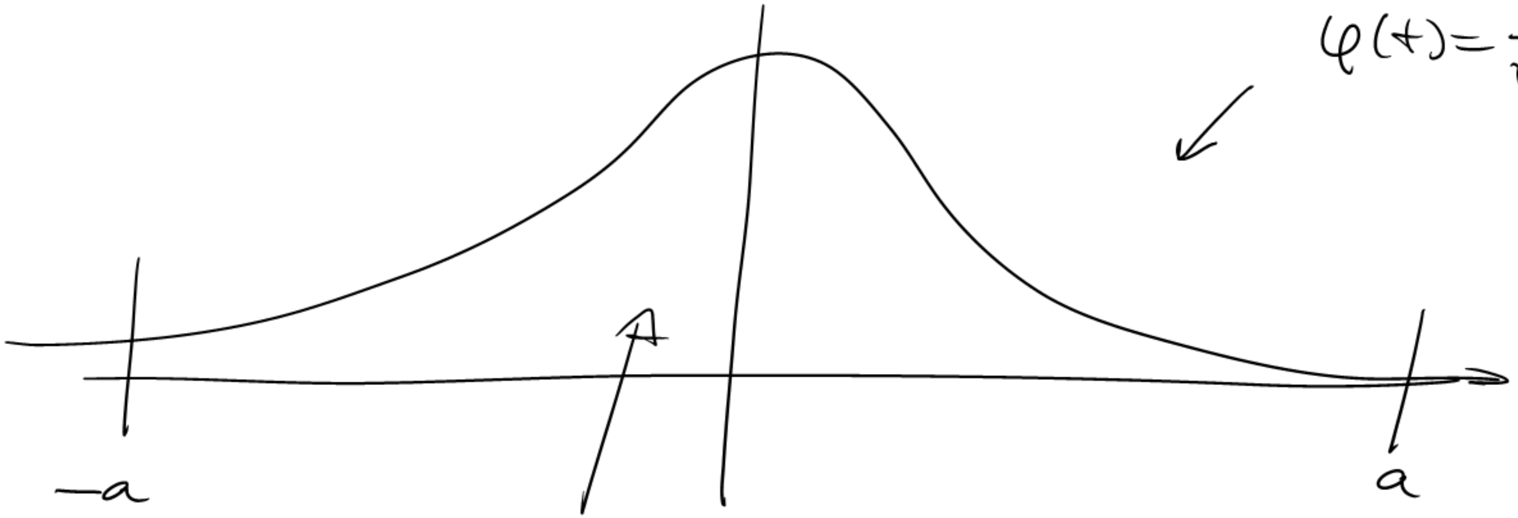
$A \subseteq D$

$$= \frac{\mathrm{vol}(A)}{\mathrm{vol}(D)} \quad \square$$

(iii) Wegen Symmetrie ist für $\vec{V} \notin D$, der Punkt $(1,1) - \vec{V} \in D$ und gleichverteilt auf D

(c)

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

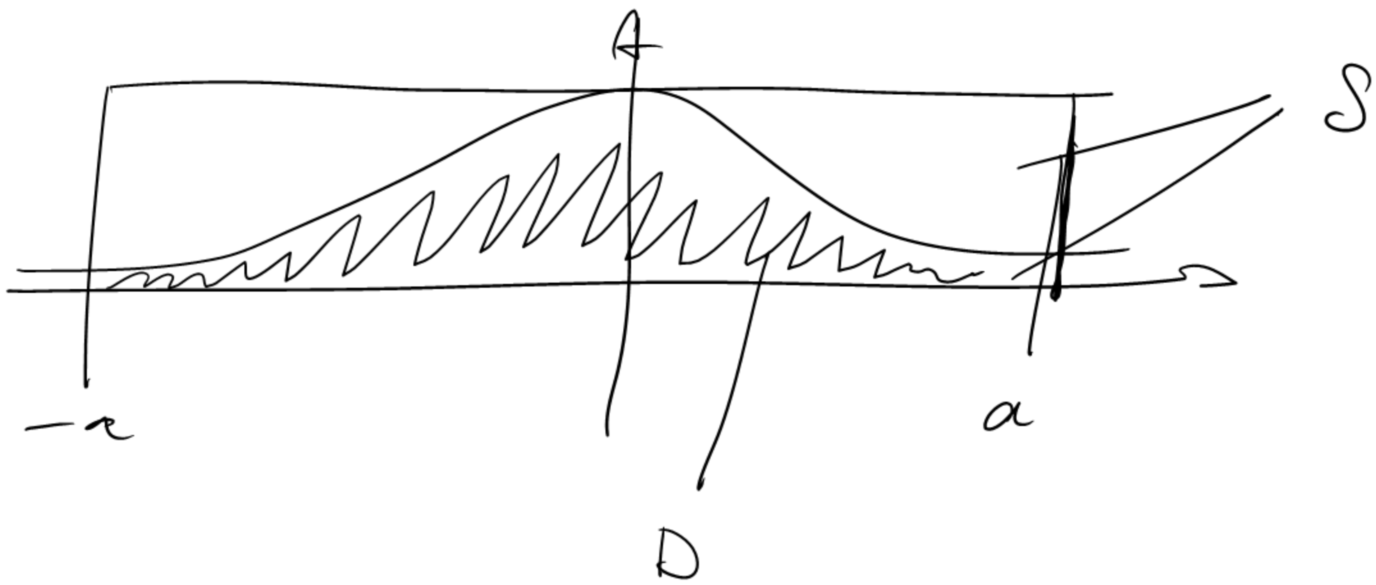


- Masse zwischen $-a$ und a: $\Phi(a) - \Phi(-a)$
- Soll mindestens 0.999 betragen:

$$\rightsquigarrow \quad \Phi(a) - \Phi(-a) = \Phi(a) - (1 - \Phi(a)) = 2\Phi(a) - 1 \overset{!}{\geq} 0.999$$

$$\Longleftrightarrow \quad a \geq \Phi^{-1}(0.9995) \approx 3.29$$

Idee wie in (b) Zufällige Punkte $\vec{v}$ in $[-a,a]\times[0,\varphi(0)]$



$\rightsquigarrow$ W'keit, dass $\vec{v}$ unter dem Graph von φ in $[-a,a]\times[0,\varphi(0)]$ fällt:

$$\frac{\Phi(a)-\Phi(-a)}{2a\cdot\varphi(0)} = \frac{0.999}{2\cdot 3.29\cdot}\sqrt{2\pi} \approx 0.380$$

$\rightsquigarrow$ Im Mittel wird jetzt also nur 38% der Samples behalten.

$\vec{V} = (V_1, V_2)$ mit $V_1 \sim \mathrm{Uni}([-a,a])$, $V_2 \sim \mathrm{Uni}([0, \varphi(0)])$ unabh.

Sei $t \in [-a,a]$.

Gesucht: $\Pr[\, V_1 \le t \mid \vec{V} \in D]$

D.h.:

$$\Pr[V_1 \le t \mid \vec{V} \in D] = \frac{\Pr[V_1 \le t \wedge \vec{V} \in D]}{\Pr[\vec{V} \in D]}$$

$$= \frac{\Pr[V_1 \le t \wedge V_2 \le \varphi(V_1)]}{\Pr[\vec{V} \in D]}$$

$$\Pr[V_1 \le t \wedge V_2 \le \varphi(V_1)]$$

$$= \int_{-a}^{t} \int_{0}^{\varphi(v_1)} \frac{1}{\varphi(0)} \cdot dv_2 \cdot \frac{1}{2a} \cdot dv_1$$

$$= \frac{1}{2a\varphi(0)} \int_{-a}^{t} \varphi(v_1)\, dv_1 = \frac{1}{2a\varphi(a)} \cdot (\Phi(t) - \Phi(-a))$$

$$\leadsto \Pr[V_1 \le t \mid \vec{V} \in D] = \frac{\frac{1}{2a\varphi(0)} \cdot (\Phi(t) - \Phi(-a))}{\frac{1}{2a\,\varphi(a)} \cdot (\Phi(a) - \Phi(-a))}$$

$$= \frac{\Phi(t) - 0.0005}{0.999}$$

$$= \Phi(t) + \underbrace{\frac{1}{999}\Phi(t) - \frac{0.5}{999}}_{\in \left[-\frac{0.5}{999}, \frac{0.5}{999}\right]} \qquad \square$$