

Kumulative Wahrscheinlichkeitsverteilungen und Funktionen von Zufallsvariablen und -Vektoren

1 Kumulative Verteilungen

Der Term “Wahrscheinlichkeitsverteilung” einer Zufallsvariable X kann manchmal etwas verwirrend sein, denn er wird auch verwendet für die (*kumulative*) *Verteilungsfunktion*

$$F_X(t) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq t\}). \quad (1)$$

Das heißt

$$F_X(t) = \begin{cases} \sum_{k \leq t} \mathbb{P}(X = k), & X \text{ ist diskret,} \\ \int_{s=-\infty}^t f_X(s) ds, & X \text{ ist kontinuierlich,} \end{cases} \quad (2)$$

wobei $f_X(s)$ die Wahrscheinlichkeitsdichte von X bezeichnet.

Aus der Definition (1) sieht man sofort, dass $\lim_{t \rightarrow -\infty} F_X(t) = 0$ und $\lim_{t \rightarrow +\infty} F_X(t) = 1$, aber auch, dass F_X *rechts-stetig* ist, also $\lim_{s \searrow t} F_X(s) = F_X(t)$. Merke, dass die nicht-strikte Ungleichheit $X(\omega) \leq t$ in (1) essentiell für diese Rechts-Stetigkeit ist.

Aus (2) schließt man, dass für kontinuierliche Zufallsvariablen, die Dichte die Ableitung der Verteilungsfunktion ist: $f_X(t) = F'_X(t)$. Manchmal schreibt man

$$\mathbb{P}(X \in A) = \int_A f_X(t) dt = \int_A dF_X(t).$$

Beispiel 1. In vielen Fällen rechnet man leichter mit der Verteilungsfunktion als mit der Dichte. Zum Beispiel, wenn X die Dichte f_1 hat, wie berechnet man die Dichte f_2 von X^2 ? Dies wird gelöst indem man erst die Verteilungsfunktion F_2 von X^2 anschaut:

$$F_2(t) = \mathbb{P}(X^2 \leq t) = \mathbb{P}(-\sqrt{t} \leq X \leq \sqrt{t}) = \int_{-\sqrt{t}}^{\sqrt{t}} f_1(s) ds.$$

Jetzt die Ableitung (mit Hilfe der Kettenregel!)

$$f_2(t) = F'_2(t) = \frac{d}{dt} \int_{-\sqrt{t}}^{\sqrt{t}} f_1(s) ds = f_1(\sqrt{t}) \frac{1}{2\sqrt{t}} - f_1(-\sqrt{t}) \frac{1}{-2\sqrt{t}} = \frac{1}{2\sqrt{t}} (f_1(\sqrt{t}) + f_1(-\sqrt{t})).$$

Dies gilt nur für $t > 0$, aber wenn $t \leq 0$, dann ist ohnehin $\mathbb{P}(X^2 \leq t) = 0$, also $f_2(t) = F'_2(t) = 0$ für $t \leq 0$.

Beispiel 2. Seien X und Y unabhängige Zufallsvariablen und $Z = \max(X, Y)$. Wir berechnen die Wahrscheinlichkeitsdichte f_Z von Z .

Erstmal die kumulative Verteilung von Z :

$$F_Z(t) = \mathbb{P}(Z \leq t) = \mathbb{P}(X \leq t \wedge Y \leq t) = \mathbb{P}(X \leq t) \mathbb{P}(Y \leq t) = F_X(t) F_Y(t).$$

Jetzt die Ableitung, mit Hilfe der Produktregel:

$$f_Z(t) = F'_Z(t) = F'_X(t) F_Y(t) + F_X(t) F'_Y(t) = f_X(t) F_Y(t) + f_Y(t) F_X(t).$$

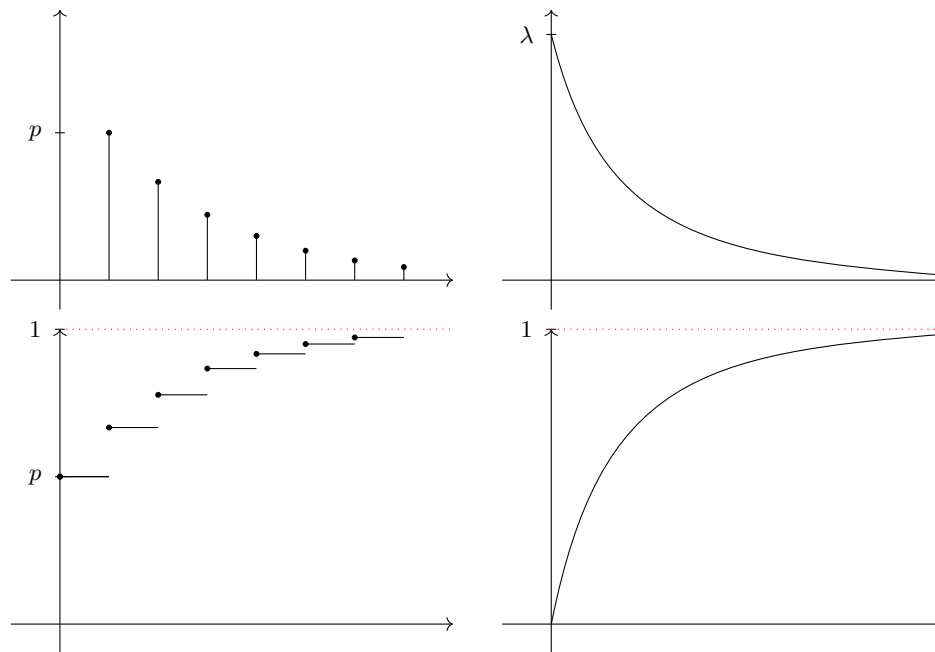


Figure 1: “Dichte” und “Verteilung” von $X \simeq \text{Geo}(p)$ links, und $X \simeq \text{Exp}(\lambda)$ rechts.

2 Die Gamma-Verteilung

Erstmal definieren wir die Gamma-Funktion

$$\Gamma(r) = \int_0^\infty y^{r-1} e^{-y} dy.$$

Also $\Gamma(1) = \int_0^\infty e^{-y} dy = 1$ und (das ist eine Übungsaufgabe) $\Gamma(r) = (r-1)!$ für ganze Zahlen $r \geq 1$. Also die Gamma-Funktion ist eine stetige Funktion deren Graphen die Punkte $(r, (r-1)!)$, $r \in \{1, 2, 3, \dots\}$, interpoliert.

Lemma 3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Proof. Übungsaufgabe. □

Definition 4. Eine Zufallsvariable $X \in [0, \infty)$ ist Gamma-verteilt ($X \simeq G(r, \lambda)$ für Parameter $r, \lambda > 0$), falls sie die folgende Dichte hat:

$$f_X(t) = \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t}.$$

Vollständige Wahrscheinlichkeit gilt, denn mit der Koordinatentransformation $u = \lambda t$ bekommen wir: $\int_0^\infty f_X(t) dt = \frac{1}{\Gamma(r)} \int_0^\infty (\lambda t)^{r-1} e^{-\lambda t} \lambda dt = \frac{1}{\Gamma(r)} \int_0^\infty u^{r-1} e^{-u} du = 1$.

Spezialfälle:

- $G(1, \lambda) = \text{Exp}(\lambda)$, die exponentielle Verteilung.
- Die $G(r, \frac{1}{2})$ -Verteilung heißt in der Statistik die χ^2 -Verteilung mit r Freiheitsgraden.

Theorem 5. Sei die Anzahl X von Ereignissen in einem Zeit-Intervall $[0, t)$ Poisson-verteilt mit Parameter $\lambda t > 0$. Dann hat die Wartezeit Y auf Ereignis Nummer r die $G(r, \lambda)$ -Verteilung.

Für Ereignis $r = 1$ ist das also die exponentielle Verteilung.

Proof. Sei t beliebig. Die Wahrscheinlichkeit, dass die Wartezeit $Y \geq t$ ist genau die Wahrscheinlichkeit, dass im Intervall $[0, t)$ höchstens $r - 1$ Ereignisse stattfanden, Also

$$F_Y(t) = 1 - \mathbb{P}(Y \geq t) = 1 - \mathbb{P}(X \leq r - 1) = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} - \frac{(\lambda t)^2}{2} e^{-\lambda t} - \dots - \frac{(\lambda t)^{r-1}}{(r-1)!} e^{-\lambda t}.$$

Jetzt die Ableitung, mit Hilfe einer Teleskopsumme:

$$\begin{aligned} f_Y(t) &= F'_Y(t) = \frac{d}{dt} \left(1 - e^{-\lambda t} - \lambda t e^{-\lambda t} - \frac{(\lambda t)^2}{2} e^{-\lambda t} - \dots - \frac{(\lambda t)^{r-1}}{(r-1)!} e^{-\lambda t} \right) \\ &= \lambda e^{-\lambda t} + (-\lambda + \lambda^2 t) e^{-\lambda t} + \left(-\lambda^2 t + \frac{\lambda^3 t^2}{2} \right) e^{-\lambda t} \dots + \left(-\frac{\lambda^{r-1} t^{r-1}}{(r-2)!} + \frac{\lambda^r t^{r-1}}{(r-1)!} \right) e^{-\lambda t} \\ &= \frac{\lambda}{(r-1)!} (\lambda t)^{r-1} e^{-\lambda t} = \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t}. \end{aligned}$$

□

3 Die Normalverteilung

Die berühmteste Verteilung ist wohl die Normalverteilung, denn seine Dichte ist die Gauss'sche Glockenkurve, die es damals bis zur Deutschen 10 Mark Banknote geschafft hat. Eine Zufallsvariable $X \in \mathbb{R}$ hat die *normale Verteilung* ($X \simeq \mathcal{N}(\mu, \sigma^2)$) falls sie die folgende Dichte hat:

$$f_X(t) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Falls $\mu = 0$ and $\sigma = 1$, so heißt diese Verteilung die *Standardnormalverteilung* $\mathcal{N}(0, 1)$.

Die Normalverteilung wird meistens verwendet um Größen von Sachen die vielen zufälligen Faktoren unterliegen. Zum Beispiel die Größe einer Person, denn die wird durch viele (genetische) Umstandsfaktoren bestimmt. Wir werden das später genauer sehen in dem Zentralen Grenzwertsatz.

Lemma 6. Sei $X \simeq \mathcal{N}(\mu, \sigma^2)$, also ist $\mathbb{E}(X) = \mu$ und $\text{Var}(X) = \sigma^2$. Zusätzlich ist $Y = (X - \mu)/\sigma$ standard normalverteilt.

Proof. Erst verwenden wir die Koordinatentransformation $y = (x - \mu)/\sigma$ die genau der Umwandlung von X in Y entspricht:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dt = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{y^2}{2}} dy.$$

Also ist die Dichte von Y gleich $f_Y(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$, also $Y \simeq \mathcal{N}(0, 1)$.

Um zu zeigen, dass $f_Y(t)$ tatsächlich eine Wahrscheinlichkeitsdichte ist, schreiben wir erstmal

$$M_n(Y) := \mathbb{E}(|Y|^n) = \frac{2}{\sqrt{2\pi}} \int_0^\infty y^n e^{-\frac{y^2}{2}} dy.$$

für das n -te Momentum.

Eine weitere Koordinatentransformation $z = y/\sqrt{2}$ und danach die Symmetrie des Integranden ergibt

$$M_0(Y) = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^\infty e^{-z^2} dz = \frac{1}{\sqrt{\pi}} \int_0^\infty e^{-z^2} dz.$$

Aber in Lemma 3 haben wir schon gezeigt, dass $\Gamma(\frac{1}{2}) = \int_0^\infty e^{-z^2} dz = \sqrt{\pi}$. Also $M_0 = 1$.

Aus der Symmetrie des Integranden schließen wir sofort, dass für ungrade $n \in \mathbb{N}$,

$$M_n(Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty y^n e^{-\frac{y^2}{2}} dy = 0.$$

Für gerade $n \geq 2$ verwenden wir partielle Integration:

$$\begin{aligned} M_n(Y) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty y^{n-1} \cdot y e^{-\frac{y^2}{2}} dy \\ &= \left[\frac{1}{\sqrt{2\pi}} y^{n-1} \cdot -e^{-\frac{y^2}{2}} \right]_{-\infty}^\infty - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty (n-1)y^{n-2} \cdot -e^{-\frac{y^2}{2}} dy \\ &= 0 + (n-1)M_{n-2}. \end{aligned}$$

Also $\text{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = M_2(Y) - M_1(Y)^2 = (2-1) \cdot M_0(Y) - 0^2 = 1$. Rücktransformation zu X ergibt $\mathbb{E}(X) = \mu$ und $\text{Var}(X) = \sigma^2$. $\square$

Bemerkung 7. Mit der oben verwendeten Koordinatentransformation können wir im Allgemeinen zeigen: falls eine kontinuierliche Zufallsvariable Erwartung μ und Varianz σ^2 hat, so ist die Zufallsvariable $Y = (X - \mu)/\sigma$ sowohl zentriert (also $\mathbb{E}(Y) = 0$) als normiert (also $\text{Var}(Y) = 1$).

Definition 8. Sei Σ eine $n \times n$ -Matrix, die positiv definit ist (also $\langle \Sigma x, x \rangle \geq 0$ für all $x \in \mathbb{R}^n$). Dann ist der Zufallsvektor $\vec{X}$ (n -dimensionell) Σ -normal verteilt falls er die folgende Wahrscheinlichkeitsdichte hat:

$$f_{\vec{X}}(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2} \langle x, \Sigma^{-1} x \rangle}, \quad x \in \mathbb{R}^n.$$

Hier ist die Matrix Σ die Kovarianzmatrix mit Einträgen $\Sigma_{i,j} = \text{Cov}[X_i, X_j]$.

4 Zufallsvektoren

Sei $\vec{X} = (X_i)_{i=1}^n$ ein n -dimensioneller Zufallsvektor. Dann sieht die Verteilungsfunktion folgendermaßen aus:

$$F_{\vec{X}}(t_1, \dots, t_n) := \mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n).$$

Die Beistriche im rechten Glied bedeuten “und”, also sollte man vielleicht $F_{\vec{X}}(t_1, \dots, t_n) := \mathbb{P}(X_1 \leq t_1 \wedge \dots \wedge X_n \leq t_n)$ schreiben, aber die Beistriche sind ganz üblich.

Beispiel 9. Wir nehmen einen kontinuierlichen Zufallsvektor $\vec{X}$ mit zwei Komponenten, und ein Block $B = [x_1, x_2] \times [y_1, y_2]$, dann gilt

$$\mathbb{P}(\vec{X} \in B) = F_{\vec{X}}(x_2, y_2) - F_{\vec{X}}(x_2, y_1) - F_{\vec{X}}(x_1, y_2) + F_{\vec{X}}(x_1, y_1),$$

wie man leicht ausrechnet mit Hilfe des Prinzips von Inklusion und Exklusion.

Beispiele von mehrdimensionellen Verteilungen von Zufallsvektoren sind die mehrdimensionelle Normalverteilung (siehe Definition 8) und auch die *multinomiale Verteilung* $\text{Mult}(n, p_1, \dots, p_d)$ die die Wahrscheinlichkeit in n Experimenten genau k_i Exemplare von Typ $i \in \{1, 2, \dots, d\}$ zu ziehen, wobei die Erfolgswahrscheinlichkeit von Typ i gleich p_i ist, anzeigt. Also

$$\mathbb{P}(X_i = k_i) = \binom{n}{k_1, \dots, k_d} p_1^{k_1} \dots p_d^{k_d}, \quad \text{mit} \quad \sum_{i=1}^d p_i = 1, \quad \sum_{i=1}^d k_i = n, \quad \binom{n}{k_1, \dots, k_d} = \frac{n!}{k_1! \dots k_d!}.$$

Die k -te Marginalverteilung eines Zufallsvektors $\vec{X} \in \mathbb{R}^d$ ist

$$F_k(t_k) := \mathbb{P}(X_k \leq t_k) = \lim_{t_1 \rightarrow \infty} \dots \lim_{t_{k-1} \rightarrow \infty} \lim_{t_{k+1} \rightarrow \infty} \dots \lim_{t_n \rightarrow \infty} F_X \left(\underbrace{t_1, \dots, t_{k-1}}_{\text{gegen } \infty}, t_k, \underbrace{t_{k+1}, \dots, t_n}_{\text{gegen } \infty} \right).$$

Die Komponenten eines Zufallsvektors sind **unabhängig** dann und nur dann als

$$F_X(t_1, \dots, t_n) = F_1(t_1) \dots F_n(t_n).$$

5 Funktionen von Zufallsvariablen und -Vektoren

In diesem Absatz versuchen wir Wahrscheinlichkeitsdichten von transformierten Zufallsvektoren zu finden. Das hat Anwendungen, wie zum Beispiel das Finden der Dichte von $X_1 + X_2$ und $X_1 X_2$, auch wenn X_1 und X_2 nicht unabhängig sind.

Sei $\vec{X}$ ein Zufallsvektor der nur Werte in einer Teilmenge U von $\mathbb{R}^d$ annimmt, und $\psi : U \rightarrow \mathbb{R}^d$ eine nicht zu üble Funktion. Der nächste Satz besagt dann was die Dichte von $\psi(X)$ ist.

Theorem 10. Sei $\vec{X} : \Omega \rightarrow U \subset \mathbb{R}^d$ ein Zufallsvektor und $\psi : U \rightarrow \mathbb{R}^d$ eine injektive, stetig differenzierbare Funktion, so dass $\det D\psi(x) \neq 0$ für alle $x \in U$. Dann hat $\vec{Y} = \psi(\vec{X})$ die Wahrscheinlichkeitsdichte

$$f_Y(y) = \mathbf{1}_{\psi(U)}(y) \cdot f_{\vec{X}}(\psi^{-1}(x)) \cdot |\det D\psi^{-1}(y)|,$$

wobei $f_{\vec{X}}$ die Dichte von $\vec{X}$ bezeichnet.

Proof. Weil $\vec{X}$ nur Werte in U annimmt, kann Y nur Werte in $\psi(U)$ annehmen. Das erklärt die Anwesenheit der Indikatorfunktion $\mathbf{1}_{\psi(U)}$.

Sei nun A irgendeine (messbare) Menge in $\psi(U)$, dann gilt

$$\mathbb{P}(Y \in A) = \mathbb{P}(\psi(X) \in A) = \mathbb{P}(X \in \psi^{-1}(A) \cap U) = \int_{\psi^{-1}(A) \cap U} f_X(x) dx.$$

Jetzt verwenden wir die Koordinatentransformation $y = \psi(x)$. Die mehrdimensionelle Transformationsformel für Integrale ergibt dann:

$$\mathbb{P}(Y \in A) = \int_{\psi(\psi^{-1}(A) \cap U)} f_{\vec{X}}(\psi^{-1}(y)) \cdot |\det D\psi^{-1}(y)| dy$$

Da $\psi(\psi^{-1}(A) \cap U) = A \cap \psi(U)$, erhalten wir

$$\mathbb{P}(Y \in A) = \int_{A \cap \psi(U)} f_{\vec{X}}(\psi^{-1}(y)) \cdot |\det D\psi^{-1}(y)| dy = \int_A \mathbf{1}_{\psi(U)} \cdot f_{\vec{X}}(\psi^{-1}(y)) \cdot |\det D\psi^{-1}(y)| dy.$$

Die Menge A war beliebig (messbar), also muss die Wahrscheinlichkeitsdichte von $\vec{Y}$ gleich $\mathbf{1}_{\psi(U)} \cdot f_{\vec{X}}(\psi^{-1}(y)) \cdot |\det D\psi^{-1}(y)|$ sein, wie behauptet. $\square$

Wir haben die folgende Liste von Folgerungen (eigentlich Anwendungen):

1. Hat die Zufallsvariable X auf $U = [0, \infty)$ die Dichte $f_X(t)$ und $\alpha \neq 0$, so hat $Y := X^\alpha$ die Dichte

$$f_Y(y) = \frac{1}{|\alpha|} y^{\frac{1}{\alpha}-1} f_X(y^{1/\alpha}), \quad y \geq 0,$$

und $f_Y(y) = 0$ für $y < 0$. Hier nehmen wir $\psi : [0, \infty) \rightarrow [0, \infty)$, $x \mapsto x^\alpha$. Dass $|\det D\psi(0)| = \infty$ für $\alpha < 1$ macht nichts, denn die 0 ist nur ein einziger Punkt mit Maß 0.

2. Die Dichte von $Y = aX + b$ für eine Zufallsvariable $X \in \mathbb{R}$ ist

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Hier nehmen wir $\psi : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto ax + b$.

3. Sei A eine invertierbare $d \times d$ -Matrix und $\vec{X}$ ist ein Zufallsvektor auf $\mathbb{R}^d$ mit Dichte $f_{\vec{X}}$, so hat $Y = A\vec{X}$ die Dichte

$$f_Y(y) = f_{\vec{X}}(A^{-1}y) \cdot |\det A^{-1}|.$$

4. Sei $\vec{X} \in \mathbb{R}^2$ ein Zufallsvektor mit Dichte $f_{\vec{X}}$ und Komponenten X_1 und X_2 . Dann hat die Zufallsvariable $Y := X_1 + X_2$ die Dichte

$$f_Y(y) = \int_{-\infty}^{\infty} f_{\vec{X}}\left(\begin{smallmatrix} y-x \\ x \end{smallmatrix}\right) dx. \quad (3)$$

Im Falle, dass X_1 und X_2 unabhängig sind, ist $f_{\vec{X}} = f_{X_1} \cdot f_{X_2}$ und finden wir die Dichte $f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y-x) \cdot f_{X_2}(x) dx$. Dies heißt die Konvolution von den Dichten f_{X_1} und f_{X_2} .

Um (3) zu zeigen nehmen wir die Matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Dann ist Y die erste Komponente von $A\vec{X}$. Wir haben $U = \mathbb{R}^2$, $A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ und $|\det A^{-1}| = 1$. Also $A\vec{X}$ hat Dichte $f_{\vec{X}}\left(A^{-1}\begin{pmatrix} y \\ x_2 \end{pmatrix}\right) = f_{\vec{X}}\begin{pmatrix} y-x_2 \\ x_2 \end{pmatrix}$.

Um die Dichte des ersten Komponenten zu erreichen integrieren wir über x_2 . Wenn wir $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ umbenennen zu $\begin{pmatrix} y \\ x \end{pmatrix}$ kommt (3) heraus.

5. Sei $\vec{X} \in \mathbb{R}^2$ ein Zufallsvektor mit Dichte $f_{\vec{X}}$ und Komponenten X_1 und X_2 . Dann hat die Zufallsvariable $Y := \frac{X_1}{X_2}$ die Dichte

$$f_Y(y) = \int_{-\infty}^{\infty} f_{\vec{X}}\left(\begin{pmatrix} xy \\ x \end{pmatrix}\right) \cdot |x| dx. \quad (4)$$

Die Methode ist wie beim vorherigen Fall, aber nun mit $U = \{(x_1, x_2) : x_1 \in \mathbb{R}, x_2 \in \mathbb{R} \setminus \{0\}\}$ und $\psi : U \rightarrow \mathbb{R}^2, (x_1, x_2) \mapsto (\frac{x_1}{x_2}, x_2)$. Dann ist $\psi^{-1}(y_1, y_2) = (y_1 y_2, y_2)$ und $\det D\psi^{-1}(y_1, y_2) = |\det \begin{pmatrix} y_2 & y_1 \\ 0 & 1 \end{pmatrix}| = |y_2|$. Schließlich integrieren wir über y_2 um die erste Komponente zu berechnen. Umbenennung $(y_1, y_2) = (y, x)$ ergibt (4).