Wahrscheinlichkeitstheorie

Dr. C.J. Luchsinger

4 Erwartungswerte

4.0 Ein paar Vorbemerkungen

1. Wir haben in Kapitel 2 nach Lemma 2.11 einen kurzen Abstecher gemacht, in dem wir eine Erweiterung von $(\Omega, \mathcal{A}, P)$ und einem darauf definierten X vorgenommen haben, um auch Werte $\pm\infty$ als Werte einer Zufallsgrösse zuzulassen. Wir brauchen dies in der Vlsg AS. Falls eine Zufallsgrösse X mit Wahrscheinlichkeit > 0 den Wert $+\infty$ annimmt (und beispielsweise nach unten beschränkt ist), dann definieren wir den Erwartungswert als $+\infty$. Es ist aber bereits in der Vlsg WTS darauf hingewiesen worden, dass eine Zufallsgrösse durchaus sowohl mit Wahrscheinlichkeit 1 endliche Werte annehmen kann, aber trotzdem keinen endlichen Erwartungswert hat. Dazu geben Sie im Must-Teil auf Blatt 10 ein einfaches Beispiel (benutzen Sie dazu den Erwartungswertsbegriff aus der WTS).

2. In der WTS haben wir in WTS-Definition 3.1 den Erwartungswert $E[X]$ einer diskreten und (absolut) stetigen Zufallsgrösse X definiert als

$$E[X] := \begin{cases} \sum_{x_i} x_i P[X = x_i] & \text{falls } X \text{ diskret} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{falls } X \text{ (absolut) stetig.} \end{cases}$$

Weiters definierten wir mit $g(x)$ eine Borel-Funktion von $\mathbb{R}$ nach $\mathbb{R}$:

$$E[g(X)] = \begin{cases} \sum_{x_i} g(x_i) P[X = x_i] & \text{falls } X \text{ diskret} \\ \int_{-\infty}^{\infty} g(x) f(x) dx & \text{falls } X \text{ (absolut) stetig.} \end{cases}$$

Diese Definitionen gelten, falls die Summe bzw. das Integral existiert. Dabei wird jeweils über den gesamten Wertebereich der Zufallsgrösse summiert respektive integriert.

Diese "Definitionen" reichen für eine einführende Vorlesung und Serviceveranstaltungen. Sie sind auch anschaulich wegen des bekannten Durchschnitts und der physikalischen

Interpretation als Schwerpunkt. Diese beiden Definitionen sind aber nicht ganz unproblematisch, wenn einfach so parallel hingeschrieben. Die erste ist ein Spezialfall der zweiten Definition und es ist nicht sofort ersichtlich, dass dies nicht auf einmal zu 2 verschiedenen Erwartungswerten führen könnte: je nachdem, ob man in einer konkreten Situation die erste oder die zweite Definition benutzt ($E[X^2] = \int x^2 f_X(x)dx$ oder $E[X^2] = \int x f_{X^2}(x)dx$). Das Problem obiger Definitionen ist aber insbesondere, dass wir in der modernen Stochastik ein besseres Fundament brauchen; deshalb wird zur Definition von Erwartungswerten weit ausgeholt (4.1-4.3, 4.4, 4.5). Wir werden die obigen Resultate in 4.5 (also relativ spät) sauber erarbeiten.

3. Was wollen wir sinnvollerweise von einem Erwartungswert fordern (haben wir in der WTS über E gelernt)?

4.1 Erwartungswert einfacher Zufallsgrössen

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Wir haben in Kapitel 2 einfache Zufallsgrössen eingeführt in dem Sinne, dass Sie nur endlich viele Werte annehmen und damit eine Darstellung in der Art

$$X = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i} \tag{D}$$

besitzen, wo $(A_i)_{i=1}^n$ eine Partition von Ω ist (die a_i's müssen *nicht* verschieden sein). Dann definieren wir erstmal

Definition 4.1 [Erwartungswert einer einfachen Zufallsgrösse] *Sei X eine einfache Zufallsgrösse mit Darstellung $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$. Dann definieren wir*

$$E[X] := \sum_{i=1}^{n} a_i P[A_i].$$

Wir orientieren uns damit bei dieser Definition durchaus am diskreten Fall von WTS-Definition 3.1: auch dort war der Erwartungwert eine *gewichtete* Summe; die Gewichte sind dabei die Wahrscheinlichkeiten $(P[A_i])$, mit denen die Werte (a_i) angenommen werden.

Sie zeigen in einer kleinen Übung, dass dies wohldefiniert in dem Sinne ist, dass der Erwartungswert *nicht* von der Darstellung (D) abhängt.

Berechnen Sie mit Hilfe von Definition 4.1 $E[\mathbf{1}_A]$ und $E[c]$, wo c eine Konstante.

Nach diesen kleinen Resultaten folgen Linearität und Monotonie von E *bei einfachen Z.G.*:

Satz 4.2 [Linearität von E bei einfachen Z.G.] *Seien X, Y beide einfache Z.G. und $a, b \in \mathbb{R}$. Dann ist auch $aX + bY$ eine einfache Z.G. und es gilt:*

$$E[aX + bY] = aE[X] + bE[Y].$$

Beweis von Satz 4.2:

Wie folgt jetzt hieraus die Monotonie von E bei einfachen Z.G.?

4.2 Erwartungswert nichtnegativer Zufallsgrössen

In Lemma 2.13 haben wir gezeigt, dass wir für jede Z.G. X mit $X \geq 0$ eine Folge von einfachen Zufallsgrössen $0 \leq X_1 \leq X_2 \ldots$ haben, sodass $X_n(\omega) \uparrow X(\omega)$ punktweise. Es ist dann naheliegend, den Erwartungswert nichtnegativer Zufallsgrössen als Limes von Erwartungswerten ebensolcher Z.G. zu definieren:

Definition 4.3 [Erwartungswert nichtnegativer Zufallsgrössen] *Sei X eine nichtnegative Zufallsgrösse. Sei X_n eine monotone, nichtnegative Folge von einfachen Zufallsgrössen, sodass $X_n \uparrow X$. Dann definieren wir*

$$E[X] := \lim_{n\to\infty} E[X_n] \leq \infty.$$

Wegen der Monotonie von E für einfache Z.G. haben wir auch $E[X_1] \leq E[X_2] \leq \ldots$. Damit existiert der Limes auf jeden Fall (monotone Folgen konvergieren immer eigentlich oder uneigentlich). Der Limes kann aber unendlich sein.

Wir müssen noch die technisch relevante Erwartung überprüfen, dass obiger $E[X]$ nicht von der gewählten Folge abhängt; dies ist der Fall:

Lemma 4.4 [E ist unabhängig von der approximierenden Folge] *Seien (X_n) und $(\tilde{X}_k)$ beides Folgen von monotonen, nichtnegativen, einfachen Z.G., welche beide gegen X konvergieren. Dann gilt auch*

$$\lim_{n\to\infty} E[X_n] = \lim_{k\to\infty} E[\tilde{X}_k].$$

Streng mathematisch gesehen, durfte man in Definition 4.3 nicht das gleiche Zeichen "E" für den Erwartungswert benutzen wie in Definition 4.1. Aber wegen Lemma 4.4 dürfen wir bei einfachen, nichtnegativen Zufallsgrössen X diese auch durch die "Folge" von $X_n := X$ für alle n "approximieren" und erhalten damit den gleichen Wert in Definition 4.1 wie in Definition 4.3. In 4.3 wird diese mathematische Raffinesse noch in den negativen Teil fortgesetzt - es wird also erlaubt sein, hier immer das gleiche Symbol E zu benutzen!

Beweis von Lemma 4.4:

Genau wie in Teil 4.1 werden wir auch hier - jetzt für eine grössere Klasse von Z.G. - die wichtigsten Eigenschaften beweisen:

Satz 4.5 [Linearität von E bei nichtnegativen Zufallsgrössen] *Seien X, Y je nichtnegative Z.G. und $a, b \in \mathbb{R}_+$. Dann gilt*

$$E[aX + bY] = aE[X] + bE[Y].$$

Beweis Satz 4.5:

Auch hier wollen wir noch auf die Monotonie von E bei nichtnegativen Z.G. schliessen:

Es folgen jetzt noch ein paar Resultate vor 4.3, welche thematisch eher zu Kapitel 5 ($n \to \infty$) gehören, beweistechnisch aber jetzt Sinn machen:

Der folgende Satz ist der erste in dieser Vlsg, welcher die allgemeine Frage in WT und Analysis behandelt:

"wann darf ich Limesbildung und Integration vertauschen?".

Satz 4.6 [Lemma von Fatou] *Sei (X_n) eine nichtnegative Folge von Zufallsgrössen. Dann gilt:*

$$E[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} E[X_n].$$

Geben Sie eine Situation an, in der obige Ungleichung strikt gilt:

In den Übungen müssen Sie im Honours-Programm noch eine Erweiterung hin zum $\limsup$ beweisen.

Beweis von Satz 4.6:

Satz 4.7 [Satz über monotone Konvergenz von Beppo Levi (1875-1961)] *Sei (X_n) eine monotone, nichtnegative Folge von Zufallsgrössen mit $X_n \uparrow X$. Dann gilt:*

$$\lim_{n\to\infty} E[X_n] = E[\lim_{n\to\infty} X_n] = E[X].$$

Ein analoger Satz gilt auch in der Analysis; vgl zB Satz 1 in §9 in Forster Analysis III.

Beweis von Satz 4.7:

Beachten Sie bitte, dass unser Beispiel nach dem Lemma von Fatou *nicht* im Widerspruch zu Satz 4.7 steht!

In den Vlsg'en WTS, AS und SM haben wir bereits mehrfach die Umformung

$$E[\sum_{k=1}^{\infty} Y_k] = \sum_{k=1}^{\infty} E[Y_k]$$

gemacht. Ohne Einschränkungen (diese waren in WTS, AS und SM immer gegeben) gilt diese Formel nicht; hingegen können wir jetzt schon beweisen:

Satz 4.8 [Partialsummen und Vertauschung von Limesbildung und Integration] *Seien $Y_k \geq 0$ und $\sum_{k=1}^{\infty} Y_k(\omega) < \infty$ für alle $\omega \in \Omega$. Dann gilt:*

$$E[\sum_{k=1}^{\infty} Y_k] = \sum_{k=1}^{\infty} E[Y_k].$$

Beweis Satz 4.8:

4.3 Erwartungswert von Zufallsgrössen (Integrierbare Zufallsgrössen)

Wir kommen hiermit zum letzten Erweiterungsschritt nach 4.1 und 4.2: wir werden jetzt auch die negativen Zufallsgrössen einbeziehen. Gegenüber 4.2 werden wir uns aber auch leicht einschränken, da wir Endlichkeit der involvierten Grössen fordern.

Wir repetieren aus Kapitel 2: $X^+ := \max\{X, 0\}, X^- := -\min\{X, 0\}$ (*beide* nichtnegativ!) und damit $X = X^+ - X^-$ und $|X| = X^+ + X^-$.

Definition 4.9 [Integrierbar, Erwartungswert, L^1] *Sei X eine Z.G.. Dann definieren wir*

a) X ist integrierbar, wenn $E[|X|] < \infty$.

b) Falls X integrierbar ist, definieren wir den Erwartungswert von X als

$$E[X] = E[X^+] - E[X^-].$$

c) Wir bezeichnen mit L^1 die Menge der integrierbaren Zufallsgrössen.

Kleine Betrachtungen zu Definition 4.9:

Satz 4.10 [Linearität von E] *Seien $X, Y \in L^1$ und $a, b \in \mathbb{R}$, dann gilt*

$$aX + bY \in L^1$$

und

$$E[aX + bY] = aE[X] + bE[Y].$$

Beweis Satz 4.10:

In kleinen Übungsaufgaben zeigen Sie noch für $X \in L^1$, dass dann $|E[X]| \leq E[|X|]$ und falls $X \leq Y \in L^1$, dann $E[X] \leq E[Y]$.

Nach Lemma von Fatou (Satz 4.6), Satz über die monotone Konvergenz (Satz 4.7), folgt hiermit ein dritter Satz über die Frage, wann Limesbildung und Integration (Summation, Erwartungswertbildung) vertauscht werden darf:

Satz 4.11 [Satz über majorisierte Konvergenz von Henri Lebesgue (1875-1941)] *Seien $X_1, X_2, \ldots$ und X integrierbare Z.G. derart, dass für alle $\omega \in \Omega$ gilt $X_n(\omega) \to X(\omega)$. Wir fordern weiter, dass eine Z.G. Y derart existiert, dass $Y \in L^1$ und $|X_n| \leq Y$ für alle n. Dann gilt:*

$$\lim_{n\to\infty} E[X_n] = E[X].$$

Ein analoger Satz gilt auch in der Analysis; vgl zB Satz 2 in §9 in Forster Analysis III.

Beweis Satz 4.11:

Vermeintliche Gegenbeispiele und Beispiele:

4.4 Integration bezüglich Verteilungsfunktionen

Die Integration bezüglich Verteilungsfunktionen ist ein Spezialfall des Riemann-Stieltjesschen Integrals (noch allgemeiner wäre das Lebesgue-Stieltjessche Integral nach Thomas Jan Stieltjes (1856-1894)). Als Sie in der einführenden Vorlesung über Analysis das Riemann-Integral kennengelernt haben, haben Sie wohl unschwer festgestellt, dass die Gewichtung einer Funktion $f(x)$ auf der x-Achse gleichmässig geschah: wenn Sie eine Treppenfunktion ψ integriert haben, so haben Sie kleine Rechtecke summiert:

$$\int_a^b \psi(x)dx := \sum_{k=1}^{n} c_k(x_k - x_{k-1}).$$

(Eine Funktion f ist demnach genau dann Riemann-integrierbar, wenn man ein Ober- und Unterintegral von Treppenfunktionen beliebig nahe zusammenführen kann.)

Eine physikalisch sinnvolle Erweiterung dieses Integral-Begriffes geht dann in die Richtung, dass man nicht mehr einfach gleichmässig (uniform, gleichgewichtet, Distanz $(x_k - x_{k-1})$) die Funktionswerte summiert, sondern eine gewichtete Summe von Funktionswerten nimmt: damit gelangen wir zum Riemann-Stieltjesschen Integral.

Definition 4.12 [Riemann-Stieltjessches Integral] *Seien f und α zwei reellwertige Funktionen auf $[a,b]$. Sei $Z := \{x_0, \ldots, x_n\}$ eine Zerlegung von $[a,b]$ und $\xi := \{\xi_1, \ldots, \xi_n\}$ ein zugehöriger Zwischenvektor, so heisst*

$$S_\alpha(f, Z, \xi) := \sum_{k=1}^{n} f(\xi_k)[\alpha(x_k) - \alpha(x_{k-1})]$$

eine Riemann-Stieltjessche Summe für f bezüglich α. Wenn für jede Folge von Zerlegungen, deren Feinheit gegen Null konvergiert, diese Summe konvergiert (die Grenzwerte fallen alle zusammen (!)), so sagen wir, dass f bezüglich α RS-integrierbar ist (Riemann-Stietjes-integriebar). Man schreibt dafür

$$\int_a^b f(x)d\alpha(x), \int_a^b f d\alpha(x) \text{ oder } \int_a^b f d\alpha.$$

Mit $\alpha(x) = x$ haben wir unser bekanntes Riemann-Integral! Für die weiteren Berechnungen zentral ist dann

Satz 4.13 [Verbindung Riemann-Integral und RS-Integral] *Ist die Funktion f Riemann-integrierbar und existiert die Ableitung α' auf $[a, b]$, so existiert*

$$\int_a^b f d\alpha$$

und es gilt

$$\text{RS}-\int_a^b f d\alpha = \text{R}-\int_a^b f\alpha' dx.$$

PhysikerInnen können bei solchen Formeln (vgl. Substitutionsregel der Integrationsrechnung) der Versuchung eines ausgeklügelten Differentialkalküls nicht widerstehen, um diese Formel zu begründen; dieses geht folgendermassen:

Solange man sich bewusst ist, dass dies noch kein Beweis obiger Formel ist, ist es OK zum Finden von Lösungen (ähnlich wie beim Lösen von DGL und PDE's). Aber es ist kein Beweis; diesen findet man zB in Heusser I: Kapitel XI.

Nach dieser Auflistung von Definition und Resultaten über das Riemann-Stieltjessche Integral aus der Analysis, wollen wir jetzt die Integration bezüglich Verteilungsfunktionen einführen. Dabei werden wir zwar sehen, dass diese Erwartungswerte eben Riemann-Stieltjessche Integrale sind, hingegen kann man die ganze Theorie auch ohne Kenntnisse des Riemann-Stieltjesschen Integrals verstehen und einführen. Genau so werden wir das - abgesehen von Querverweisen - auch machen.

Es sei - ebenfalls vorbereitend - darauf hingewiesen, dass im ganzen bisherigen Kapitel 4 bei jedem Erwartungswert

$$E[X]$$

einer Zufallsgrösse X auch eine Wahrscheinlichkeit P im Spiel ist (Definition 4.1 und darauf basierende Erweiterungen). Deshalb indexiert man den Erwartungswert manchmal, sobald die dazugehörige Wahrscheinlichkeit P eine Rolle spielt in der Art

$$E_P[X].$$

Nebenbemerkung: Zum Beispiel in der Finanzmathematik werden Sie nicht nur das "normale", aus historischen Daten geschätzte, P für die Entwicklung eines stochastischen Prozesses (zum Beispiel Aktienkurs S_t) haben, sondern auch ein sogenanntes "Risk-Neutral-Measure" Q. Sie haben dann - was Anfänger sehr stark verwirrt - ein $E_P[S_t]$ und ein $E_Q[S_t]$.

Falls wir jetzt eine Verteilungsfunktion F auf $\mathbb{R}$ haben, dann gibt es wegen Satz 1.17 ein eindeutiges P auf $\mathcal{B}(\mathbb{R})$ derart, dass $F_P = F$. Es gilt dann

$$P[(a, b]] = F(b) - F(a). \qquad \text{(Gewicht)}$$

Wir haben beim Exkurs über das Riemann-Stieltjessche Integral gesagt, dass man mit dem Riemann-Stieltjesschen Integral eine unterschiedliche Gewichtung bei der Integration erreicht. Genau dies machen wir (versteckt) mit Hilfe von (Gewicht) in nachfolgender Definition. Die nachfolgenden Funktionen sollen alle Borel-messbar sein.

Definition 4.14 [Integral bezüglich einer Verteilungsfunktion] *Sei F eine Verteilungsfunktion auf $\mathbb{R}$. Dann definieren wir*

a) für eine nichtnegative Funktion g das Integral von g bezüglich F als

$$\int_{\mathbb{R}} g(x)dF(x) := E_F[g] \leq \infty,$$

wobei wir (vgl. oben) $E_F[g]$ und $E_P[g]$ simultan verwenden. $E_F[g]$ ist ein uns bekannter Ausdruck: es ist der Erwartungswert der Zufallsgrösse g auf dem Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_F)$ (vgl Teil 4.2). Man benutzt an dieser Stelle üblicherweise ein kleines g für eine Zufallsgrösse (!) wegen späterer Formeln!

b) Analog zu 4.3 definieren wir: eine Funktion g ist integrierbar bezüglich F, falls

$$\int_{\mathbb{R}} |g(x)|dF(x) < \infty.$$

Wir definieren in dem Fall das Integral von g bezüglich F als

$$\int_{\mathbb{R}} g(x)dF(x) := \int_{\mathbb{R}} g^+(x)dF(x) - \int_{\mathbb{R}} g^-(x)dF(x).$$

Da Sie diese Integrationsform noch nicht kennen, als Erklärungsversuch ein Resultat aus der Zukunft: wir werden später sehen, dass $E[X] = \int_{-\infty}^{\infty} x \, dF_X(x)$. Angenommen, X ist eine $\mathcal{N}(\mu, \sigma^2)$-Zufallsgrösse. Dann müsste ja gelten, dass $E[X] = \int_{-\infty}^{\infty} x \, dF_X(x) = \mu$. Wie kann aber ein Integral über ganz $\mathbb{R}$ (!) von x einen endlichen Wert liefern, ja überhaupt definiert sein? Da hilft eben die ungleiche Gewichtung durch dF, welche wir im Riemann-Stieltjesschen Integral haben:

Wir haben in Definition 4.14 das Integral bezüglich einer Verteilungsfunktion *definiert* als etwas uns wohl bekanntes, als einen Erwartungswert. Damit können wir jetzt aber alle Resultate aus 4.2 und 4.3 importieren - sie müssen auch für dieses Integral gelten. Wir repetieren die wichtigsten Resultate summarisch und übersetzen sie gleich in die neue Sprache:

* $g \equiv c \Rightarrow \int g dF = c$ (nicht so bei Riemann-Integral)

* $B \in \mathcal{B}(\mathbb{R}) \Rightarrow \int 1_B dF = P_F(B)$

* g, h je $\geq 0; a, b \in \mathbb{R}_+$ oder g, h integrierbar und $a, b \in \mathbb{R}$ dann (**Linearität**)

$$\int (ag + bh) dF = a \int g dF + b \int h dF$$

* **Monotonie**: $0 \leq g \leq h$ oder g, h integrierbar und $g \leq h \Rightarrow \int g dF \leq \int h dF$

* **Fatou**: $g_n \geq 0$ für alle n, dann $\int \liminf_n g_n dF \leq \liminf_n \int g_n dF$

* **Monotone Konvergenz**: $0 \leq g_1 \leq g_2 \ldots \leq g_n \uparrow g$ für alle x, dann $\int g_n dF \uparrow \int g dF$.

* **Majorisierte Konvergenz**: $g_n \to g$ für alle x und existiert h integrierbar, sodass $|g_n| \leq h$ für alle n, dann $\int g_n dF \to \int g dF$.

Wo stehen wir? Wir können (ausser mit Hilfe von unbewiesenen Resultaten aus WTS) zum Beispiel nicht mal einen Erwartungswert einer absolut-stetigen Zufallsgrösse berechnen (ausser wir approximieren ihn mit Hilfe einer monoton wachsenden Folge von Erwartungswerten von einfachen Zufallsgrössen - viel Spass!). Wir werden jetzt zuerst untersuchen, wie ein Integral bezüglich einer Verteilungsfunktion aussieht, wenn die Verteilungsfunktion diskret bzw absolut-stetig ist. Damit können wir dann in 4.5 endlich die Formeln aus WTS (WT-Kapitel 4.0) herleiten.

Integrale bezüglich *diskreter* Verteilungsfunktionen sind Summen:

Satz 4.15 [Integral bezüglich einer diskreten Verteilungsfunktion] *Sei* $F(t) = \sum p_i \mathbf{1}(t_i \leq t)$, *dann gilt für alle nichtnegativen* g

$$\int g dF = \sum_i p_i g(t_i). \tag{4.1}$$

Beweis Satz 4.15:

Korollar 4.16 [integrierbares g bei diskretem F] g *ist integrierbar bzgl* F *genau dann wenn*

$$\sum_i p_i |g(t_i)| < \infty.$$

In dem Fall gilt (4.1).

Integrale bezüglich *absolut-stetiger* Verteilungsfunktionen sind Riemann-Integrale:

Satz 4.17 [Integral bezüglich einer absolut-stetigen Verteilungsfunktion] *Sei F eine absolut-stetige Verteilungsfunktion mit (stückweise) stetiger Dichtefunktion f. Sei g nichtnegativ und (stückweise) stetig. Dann gilt:*

$$\int g dF = \int_{-\infty}^{\infty} g(x)f(x)dx, \tag{4.2}$$

dabei haben wir auf der rechten Seite jetzt ein (normales) Riemann-Integral (vgl Satz 4.13).

Beweis-Skizze Satz 4.17 (Vervollständigung in den Übungen):

Korollar 4.18 [integrierbares g bei absolut-stetigem F] *Sei F eine absolut-stetige Verteilungsfunktion mit (stückweise) stetiger Dichtefunktion f. Sei g (stückweise) stetig. Dann ist g integrierbar bzgl F, genau dann wenn*

$$\int_{-\infty}^{\infty} |g(x)|f(x)dx < \infty,$$

wobei dann (4.2) gilt.

4.5 Operationelle Formeln zur Berechnung von Erwartungswerten

Wir werden in den folgenden 4 Resultaten (Satz 4.19, Korollar 4.20, Satz 4.21 und Satz 4.22), meist zuerst immer noch relativ abstrakt, die zentralen Formeln herleiten. Diese müssen danach noch explizit für die Anwendung ausgedeutscht werden (am Schluss von Teil 4.5).

Satz 4.19 [Operationelle Formel bei nichtnegativen Zufallsgrössen] *Sei* $X \geq 0$. *Dann gelten:*

$$E[X] = \int_0^\infty x dF_X(x) = \int_0^\infty [1 - F_X(y)]dy. \tag{4.3}$$

Wir haben also in der Mitte ein Integral bzgl einer Verteilungsfunktion, wobei jetzt $g(x) \equiv x$; das Integral rechts ist ein *Riemann*-Integral! Wir kennen die Formel rechts bereits aus der WTS.

Beweis von Satz 4.19

Der Spezialfall, wo X nur Werte auf $\mathbb{N}_0$ annimmt, war schon in der WTS speziell hervorgehoben worden und wird in der AS eingesetzt:

Korollar 4.20 [Operationelle Formel bei Zufallsgrössen mit Werten nur in $\mathbb{N}_0$] *Sei $P[X \in \mathbb{N}_0] = 1$. Dann gelten:*

$$E[X] = \sum_{n=0}^{\infty} nP[X = n] = \sum_{k=1}^{\infty} P[X \geq k]. \tag{4.4}$$

Beweis von Korollar 4.20

Wir werden jetzt auch negative Werte erlauben und erhalten dazu

Satz 4.21 [Operationelle Formel bei $X \in L^1$] *Sei $X \in L^1$. Dann gilt*

$$E[X] = \int_{-\infty}^{\infty} x dF_X(x). \tag{4.5}$$

Beweis von Satz 4.21

Zum Schluss noch der Fall, wo wir nicht nur den Erwartungswert einer Zufallsgrösse X, also $E[X]$ berechnen wollen, sondern - wo g borelsch - den Erwartungswert einer transformierten Zufallsgrösse $g(x)$, also $E[g(X)]$. Dazu könnte man theoretisch einfach die Verteilungsfunktion von $g(X)$ berechnen und dann mit den bisherigen Formeln fortfahren. Dies kann schwierig werden - es ist Gott sei Dank auch nicht notwendig:

Satz 4.22 [Operationelle Formel für $E[g(X)]$] *Sei $g(X) \in L^1$ oder zumindest g nichtnegativ. Dann gilt*

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) dF_X(x). \tag{4.6}$$

Beweis von Satz 4.22

Auf Englisch nennt man obige Formel(n) zur Berechnung von $E[g(X)]$ auch the "Law of the unconscious statistician" weil die (angewandten) StatistikerInnen im Gegensatz zu den hard core MasstheoretikerInnen die Mathematik hinter obiger Formel nicht sehen: http://en.wikipedia.org/wiki/Law_of_the_unconscious_statistician.

Warum haben wir dieses $\int gdF$ überhaupt eingeführt? Es erlaubt uns eine ökonomische Beweisführung: möglichst viel (Satz 4.19, 4.21 und 4.22) wird gemeinsam für alle Verteilungsarten gezeigt. Wir haben nach Satz 1.29 (Vollständige Klassifikation der Wahrscheinlichkeiten) und mit der dortigen Bezeichnung für jede Verteilungsfunktion F eine Darstellung der Art $F = aF_d + bF_a + cF_s$. Danach kann man dann noch $a = 1$ oder $b = 1$ setzen und erhält dann die üblichen Spezialfälle.

4.6 L^p-Räume und Ungleichungen

In diesem Teil folgen wichtige Ungleichungen im Zusammenhang mit Erwartungswerten. Man kann in einer allgemeinen Masstheorie-Vorlesung auf einem höheren Niveau viele der nachfolgenden Resultate abstrakter herleiten; die wichtigsten Anwendungen sind dann in der Analysis (vgl Forster Analysis III, §10) und hier nachfolgend in der Wahrscheinlichkeitstheorie. Meist gibt es ein Pendant der folgenden Resultate mit **Erwarungswerten** in der Form von **Integralen (Riemann- und mehrdimensionale Lebesgue-Integrale)** oder **Summen**.

Definition 4.23 [**L^p-Raum**] *Sei $1 \leq p < \infty$. Dann bezeichnen wir mit L^p die Menge der Zufallsgrössen X derart, dass $E[|X|^p] < \infty$ (p-tes Moment immer noch integrierbar).*

Um uns kommende Beweise zu vereinfachen, behandeln wir noch

Lemma 4.24 [**Young's Ungleichung**] *Sei $h : \mathbb{R}_+ \to \mathbb{R}_+$ eine stetige, streng monoton wachsende Funktion mit $h(0) = 0$ und $h(\infty) = \infty$. Sei k die punktweise Inverse von h. Wir definieren $H(x) := \int_0^x h(y)dy$ und $K(x) := \int_0^x k(y)dy$. Dann gilt für alle $a, b \in \mathbb{R}_+$,*

$$ab \leq H(a) + K(b).$$

Beweis Lemma 4.24

Wir halten hier noch fest, dass natürlich weiterhin die Jensen-Ungleichung (WTS-Lemma 3.5) und die Ungleichung(en) von Bienayme-Tschebyschew (und artverwandte) (WTS-Satz 5.1) gelten; Sie beweisen in den Übungen eine Verallgemeinerung von WTS-Satz 5.1.

Satz 4.25 [Höldersche Ungleichung] *Seien $p, q > 1$ derart, dass*

$$\frac{1}{p} + \frac{1}{q} = 1,$$

mit $X \in L^p$ und $Y \in L^q$. Dann gilt $XY \in L^1$ und

$$E[|XY|] \leq E[|X|^p]^{1/p} E[|Y|^q]^{1/q}.$$

Die analogen Sätze in der Analysis sind Satz 7 in §16 in Forster Analysis I (**Summen**), Beispiel (18.5) in §18 in Forster Analysis I (**Riemann-Integral**) und Lemma 1 in §10 in Forster Analysis III (**mehrdimensionale Lebesgue-Integrale**).

Beweis Höldersche Ungleichung

Korollar 4.26 [Cauchy-Schwarz Ungleichung] *Seien $X, Y \in L^2$. Dann ist $XY \in L^1$ und*

$$E[|XY|] \leq \sqrt{E[|X|^2] E[|Y|^2]}\ .$$

Die allgemeinste Form dieses Resultats ist wohl in der Linearen Algebra die gleichnamige Ungleichung (Fischer LinAlg, 5.1.3 und 5.4.7). Die analogen Sätze in der Analysis finden sich bei der Hölderschen Ungleichung an oben genannten Stellen. Die Cauchy-Schwarz Ungleichung ist offensichtlich ein Spezialfall der Hölderschen Ungleichung wo $p = q = 2$; damit ist nichts mehr zu beweisen.

Nach diesen schönen Analogien zwischen der Analysis und der WT kommt jetzt leider ein wichtiger Unterschied: Wir haben bereits in der WTS in den Übungen (damals ohne den jetzigen Überbau - jetzt in der neuen Sprache) bewiesen, dass wenn $1 \leq r \leq s$, dann gilt $L^s \subseteq L^r$; der Beweis geht (WT hat endliches Mass!) folgendermassen:

Warum gilt das analoge Resultat nicht in der Analysis? Gegenbeispiel:

Sie untersuchen für die WT noch den Fall wo $0 < r \leq s \leq 1$ in den Übungen.

Neben obigen Inklusionen in der WT gilt sogar weitergehend, dass $E[|X|^p]^{1/p}$ wachsend ist in p, $p \geq 1$:

Korollar 4.27 [Lyapunov-Ungleichung] *Sei $1 \leq r \leq s$ und $X \in L^s$. Dann gilt*

$$E[|X|^r]^{1/r} \leq E[|X|^s]^{1/s}.$$

Beweis Korollar 4.27

Wir haben in der WTS bereits (spätestens bei der Definition der Varianz) gesehen, dass für $X \in L^2$ immer gilt

$$E[X^2] \geq E[X]^2.$$

Man kann dies jetzt auf ganz viele Arten mit Hilfe obiger Sätze beweisen bzw memorieren, wie alles?

Wir kommen jetzt zu einem unscheinbaren Resultat (Dreiecksungleichung), welches dann gewaltige Konsequenzen hat:

Satz 4.28 [Minkowski-Ungleichung] *Sei* $p \geq 1$ *und* $X, Y \in L^p$. *Dann ist auch* $X + Y \in L^p$ *und*

$$E[|X+Y|^p]^{1/p} \leq E[|X|^p]^{1/p} + E[|Y|^p]^{1/p}.$$

Die analogen Sätze in der Analysis sind Satz 8 in §16 in Forster Analysis I (**Summen**), Beispiel (18.5) in §18 in Forster Analysis I (**Riemann-Integral**) und Corollar zu Lemma 1 in §10 in Forster Analysis III (**mehrdimensionale Lebesgue-Integrale**).

Beweis Satz 4.28

Nach Wikipedia (22. Mai 2008): Minkowski (1864-1909); 1896-1902 an ETHZ, Kollege von Hurwitz, Albert Einstein war Schüler von Minkowski.

Was folgt jetzt aus Satz 4.28? Skizze weiterer Pfad in WT und Funktional-Analysis