K8 Stetige Zufallsvariablen — Theorie und Praxis

8.1 Theoretischer Hintergrund Wir haben

- (nicht abzählbare) Wahrscheinlichkeitsräume
- Meßbare Funktionen
- Zufallsvariablen
- Verteilungsfunktionen
- Dichten in $\mathbb{R}^n$
- Funktionen von Zufallsvariablen
- Unabhängigkeit
- Erwartungswerte

8.2 Und in der Praxis? Modellierung

- Was ist eine Zufallsvariable?
  z.B. Länge einer Dienstfahrt
  Niederschlag pro Tag
  Unsicherheit bei Geschwindigkeitsmessungen

- Eine Verteilungsfunktion? $F(x) = P(X \leq x)$
  z.B. Was ist die Wahrscheinlichkeit, dass die Fahrt weniger als eine Stunde dauert?
  Dass wir weniger als 2 cm Regen haben?
  Dass der Meßfehler kleiner 5 km/Stunde ist?

- Eine Dichte? $P(X \leq x) = \int_{-\infty}^{x} f(x)\mathrm{d}x$
  Keine Wahrscheinlichkeit!

- Ein Erwartungswert? $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$
  z.B. Wie lange dauert die Fahrt im Durchschnitt?
  Wieviel Regen erwarten wir?
  Sind die Messungen verzerrt?

8.3 Einige Anwendungsfragen

1. Wieviele Leute sollen in einem Aufzug erlaubt werden?

2. Wieviel Gepäck darf man im Flugzeug mitnehmen?

3. Wenn die Standards verlangen, dass 95% aller Schokoriegel mindestens soviel wiegen, wie auf der Hülle versprochen wird, wie hoch muß das Durchschnittsgewicht sein?

4. Teil A muß mit Teil B zusammenpassen. Wie groß dürfen die Variabilitäten sein?

5. Blutdruck sollte nicht zu hoch sein. Wieviele haben einen zu hohen Blutdruck?

6. Es wird angenommen, dass Aktienpreise lognormalverteilt sind. d.h. für einen Aktien Preis X_t

$$Y_t = \log X_t \sim N(\mu_t, \sigma_t^2)$$

wobei $\mu_t = \log X_0$. Was ist $P(a < X_t \leq b)$?

8.4 Erwartungswerte wozu?

- „Durchschnittliche“Werte
  Gewinne, Kosten, langfristige Summen

- Beschreibung einer Verteilung
  Mittelwert, Varianz, Schiefe, die Momente

- Schätzen der Parameter einer Verteilung
  was erwarten wir von einem Schätzverfahren?

- Bewertung von finanziellen Derivaten
  was ist eine Option wert?

8.5 Parameter

$E[X]$ der "Mittelwert" einer Verteilung (1)

Die Streuung wird mit der Varianz bemessen:

$$V[X] = E[(X - E[X])^2] \tag{2}$$

oder die Standardabweichung

$$\sigma_X = \sqrt{V[X]}$$

Eigenschaften der Varianz:

$$V[X + a] = V[X]$$

$$V[aX] = a^2 V[X]$$

$$V[X] = E[X^2] - E[X]^2$$

$g(X) = x^2$ konvex $\Rightarrow E[X^2] \geq E[X]^2$

Schiefe

$$E[\frac{(X - E[X])^3}{\sigma^3}] \tag{3}$$

und Kurtosis

$$E[\frac{(X - E[X])^4}{\sigma^4}] \tag{4}$$

gibt es auch, werden aber kaum benutzt.

8.5.1 Parameter Beispiele

Gleichverteilung $U(a, b)$

$$E[X] = \frac{b + a}{2} \quad V[X] = \frac{(b - a)^2}{12}$$

Exponentialverteilung $E(\lambda)$

$$E[X] = \frac{1}{\lambda} \quad V[X] = \frac{1}{\lambda^2}$$

Normalverteilung $N(\mu, \sigma^2)$

$$E[X] = \mu \quad V[X] = \sigma^2$$

8.6 Kovarianz und Korrelation

Seien X und Y Zufallsvariablen. Die Kovarianz von X und Y ist

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \tag{1}$$

Für X und Y unabhängig gilt

$$\text{Cov}(X, Y) = 0$$

Die Korrelation von X und Y mißt die lineare Abhängigkeit zwischen X und Y:

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \tag{2}$$

$$-1 \leq \rho_{XY} \leq 1$$

Das kann mit der Cauchy-Schwarz Ungleichung gezeigt werden.

8.6.1 Die Ungleichung von Cauchy-Schwarz

Seien X und Y Zufallsvariablen und

$$E[X^2]E[Y^2] < \infty$$

dann gilt

$$E[XY]^2 \leq E[X^2]E[Y^2] \tag{3}$$

Beweis:

$$E[(XE[Y^2] - YE[XY])^2] \geq 0$$

$$\Rightarrow E[X^2]E[Y^2]^2 - 2E[XY]^2E[Y^2] + E[Y^2]E[XY]^2 \geq 0$$

$$E[Y^2](E[X^2]E[Y^2] - E[XY]^2) \geq 0$$

Für $E[Y^2] > 0$ ist das Resultat klar. Für $E[Y^2] = 0$ müssen $P(Y = 0) = P(XY = 0) = 1$ folgen und dann das Resultat.

Setzen wir $W = X - E[X]$ und $T = Y - E[Y]$.
Cauchy-Schwarz bedeutet, dass

$$E[WT]^2 \leq E[W^2]E[T^2]$$

oder

$$|\mathrm{Cov}(X,Y)| \leq \sigma_X \sigma_Y$$

$$\Rightarrow |\rho_{XY}| \leq 1$$

8.7 Auswertung von Wahrscheinlichkeiten für $N(\mu, \sigma^2)$

$$P(a \leq X < b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x$$

Numerisch integrieren — besser Experten überlassen.

$$P(a \leq X \leq b) = P(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma})$$

$$\text{wo} \quad Z \sim N(0, 1)$$

Zum Beispiel, gegeben $X \sim N(i, 1)$ was ist $P(X < 0)$?

$P(Z < i)$	Excel	R	Data Desk
0	0.5	0.5	0.5
-1	0.1587	0.1587	0.1587
-5	2.867E-07	2.867E-07	2.867E-07
-8	6.221E-16	6.221E-16	6.22E-16
-10	7.620E-24	7.620E-24	0

Data Desk geht nur bis höchstens 18 Stellen.
Vor drei Jahren hat Excel sogar $P(Z < -5)$ falsch berechnet, aber jetzt ist sie viel besser.
Was sagt Maple bzw. Mathematica?

Nach Marsaglia
"Evaluating the Normal Distribution Function"

ist der wahre Wert für $P(Z < -10)$

$$7.61985302416053E - 24$$

und er bietet eine C Routine an, die

$$7.61985302416052E - 24$$

schafft.

Excel liefert

$$7.61985E - 24$$

und R liefert

$$7.619853E - 24$$

bzw.

$$7.6198530241605269199e - 24$$

mit

```
options(digits=20)
pnorm(-10)
```

Anwendungen — Stetige Wahrscheinlichkeiten

“Proximus ecclesiae semper vult ultimus esse.”

Warum? Nehmen wir an, dass die Fahrtzeit von der Entfernung, d, abhängt:

$$X \sim N(\mu(d), \sigma^2(d))$$

und dass man das Haus früh genug verlässt, so dass

$$P(\text{man kommt zeitlich an}) \geq 1 - \alpha$$

Aktienmärkte

Ein gängiges Modell für Aktienpreisbewegungen nimmt an, dass Log Änderungen normalverteilt sind, d.h.
Sei $X(t)$ der Preis am Tag t, gilt

$$Y(t) = \log \frac{X(t)}{X(t-1)} \sim N(\mu, \sigma^2)$$

1. Was ist die Verteilung der Änderung nach einem Jahr (nehmen wir 250 Tage pro Jahr an)?

2. Was ist die Wahrscheinlichkeit, dass eine Aktie im Wert in einem Jahr doppelt? Was ist die Wahrscheinlichkeit, dass mindestens eine Aktie aus N im Wert doppelt? Was ist die Wahrscheinlichkeit, dass genau eine Aktie aus N im Wert doppelt? Was ist die Verteilung der maximalen Verbesserung aus N Aktien?

3. Was ist die Verteilung des Werts eines Aktienpakets von N Aktien mit ursprünglichen Gewichten w_i nach n Tage?

Lebensdauer von Geräten

Die Lebensdauer von Geräten, T, kann manchmal mit einer Rayleigh Verteilung modelliert werden:

$$f(t) = \frac{t}{\sigma^2} e^{-\frac{1}{2}\frac{t^2}{\sigma^2}}$$

$$0 \leq t < \infty \qquad \sigma > 0$$

1. Was ist der Erwartungswert von T?

2. Was ist die Dichte einer Rayleighverteilten ZV gegeben $T > s$?

3. Was ist der Erwartungswert der verbleibenden Lebensdauer?

4. Die Hazardrate einer ZV T zum Zeitpunkt t wird als der Quotient der Dichte und der Wahrscheinlichkeit

$P(T > t)$ definiert. Was ist die Hazardrate einer Rayleighverteilten ZV? Was ist die Hazardrate einer exponentialverteilten ZV? Was ist die Hazardrate einer Chiquadratverteilten ZV? Wie soll man diese Hazardraten interpretieren?

5. Der Hersteller bietet einen vollen Ersatz für das Produkt, falls es innerhalb von einer Zeit s_0 nach dem Kauf nicht mehr geht. Was sind die erwarteten Kosten dieser Garantie?

6. Wenn jemand drei Geräte hintereinander benutzt, was ist die Verteilung der Gesamtlebensdauer? Gegeben, dass die zwei ersten beide unter Garantie ersetzt worden sind, aber die dritte nicht, was ist die Verteilung der Gesamtlebensdauer?

7. Gegeben einer Gesamtlebensdauer der drei Geräte von s_g, was ist die Wahrscheinlichkeit, dass keins unter Garantie ersetzt werden müßte?