Übungsblatt 8 zur Vorlesung "Einführung in die Statistik"

Olivier Warin

18. November 2012

Aufgabe 55 [Bedingte Dichte ist tatsächlich Dichte]
Seien X und Y zwei stetige Zufallsgrössen mit Dichten f_X und f_Y. Sei weiter $y_0 \in \mathbb{R}$ mit $f_Y(y_0) \neq 0$. Wir definieren, wie in Formel (3.2), eine Funktion $f_{X|Y=y_0} : \mathbb{R} \to \mathbb{R}$ durch

$$f_{X|Y=y_0}(x) \;=\; \frac{f_{X,Y}(x,y_0)}{f_Y(y_0)},$$

wobei $f_{X,Y}$ die gemeinsame Dichte von X und Y bezeichnet.

Behauptung: $f_{X|Y=y_0}$ ist eine Dichtefunktion.

Beweis: $f_{X|Y=y_0}$ ist klar integrierbar und nicht negativ. Ausserdem gilt nach Turnübung 2 in 2.4:

$$f_Y(y_0) \;=\; \int_{-\infty}^{\infty} f_{X,Y}(x,y_0)dx.$$

Wir schliessen:

$$\int_{-\infty}^{\infty} f_{X|Y=y_0}(x)dx \;=\; \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y_0)}{f_Y(y_0)}dx \;=\; \frac{1}{f_Y(y_0)}\int_{-\infty}^{\infty} f_{X,Y}(x,y_0)dx \;=\; \frac{f_Y(y_0)}{f_Y(y_0)} \;=\; 1.$$

Also ist $f_{X|Y=y_0}$ wirklich eine Dichte.

■

Aufgabe 56 [einfache Turnübungen *Cov*, *Cor*]
Seien X und Y zwei Zufallsgrössen.

a) Behauptung: Es gilt:

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])] = E[X(Y-E[Y])] = E[(X-E[X])Y] = E[XY]-E[X]E[Y].$$

Beweis: Nach Definition 3.9 a) gilt:

$$\begin{aligned} Cov(X,Y) &= E[(X-E[X])(Y-E[Y])] \;=\; E[X(Y-E[Y]) - E[X](Y-E[Y])] \\ &\overset{\text{Lem}}{\underset{\text{3.4b)}}{=}} E[X(Y-E[Y])] - E[X](E[Y]-E[Y]) \;=\; E[X(Y-E[Y])] \\ &\overset{\text{Sym-}}{\underset{\text{metrie}}{=}} E[(X-E[X])Y] \;=\; E[XY - E[X]Y] \overset{\text{Lem}}{\underset{\text{3.4b)}}{=}} E[XY]-E[X]E[Y]. \end{aligned}$$

■

b) Behauptung: Seien $a,b,c,d \in \mathbb{R}$. Dann gilt:

$$Cov(aX+b, cY+d) \;=\; acCov(X,Y).$$

Beweis: Es gilt nach Definition 3.9 a):

$$\begin{aligned} Cov(aX+b,cY+d) &= E[(aX+b-E[aX+b])(cY+d-E[cY+d])] \\ &\overset{\text{Lem}}{\underset{\text{3.4b)}}{=}} E[(aX+b-aE[X]-b)(cY+d-cE[Y]-d)] \\ &= E[ac(X-E[X])(Y-E[Y])] \overset{\text{Lem}}{\underset{\text{3.4b)}}{=}} acE[(X-E[X])(Y-E[Y])] \\ &\overset{\text{Def}}{\underset{\text{3.9a)}}{=}} acCov(X,Y). \end{aligned}$$

■

c) Seien $a, b, c, d \in \mathbb{R}$ mit $a, c \neq 0$. Wir nehmen hier weiter an, dass gilt: $V[X], V[Y] > 0$.

BEHAUPTUNG: Es gilt:

$$|Cor(aX + b, cY + d)| = |Cor(X, Y)|.$$

Beweis: Mit Definition 3.9 b), Teilaufgabe b) und Lemma 3.7 a) schliessen wir:

$$\begin{aligned} |Cor(aX + b, cY + d)| &\overset{\text{Def}}{\underset{3.9b)}{=}} \frac{|Cov(aX + b, cY + d)|}{\sqrt{V[aX + b]V[cY + d]}} \overset{b)}{=} \frac{|acCov(X, Y)|}{\sqrt{a^2V[X]c^2V[Y]}} \\ &= \frac{|Cov(X, Y)|}{\sqrt{V[X]V[Y]}} \overset{\text{Def}}{\underset{3.9b)}{=}} |Cor(X, Y)|. \end{aligned}$$

■

d) BEHAUPTUNG: Es gilt:

$$V[X + Y] = V[X] + V[Y] + 2Cov(X, Y).$$

Beweis: Es gilt:

$$\begin{aligned} V[X + Y] &\overset{\text{Lem}}{\underset{3.7b)}{=}} E[(X + Y)^2] - (E[X + Y])^2 \\ &\overset{\text{Lem}}{\underset{3.4b)}{=}} E[X^2] + 2E[XY] + E[Y^2] - (E[X])^2 - 2E[X]E[Y] - (E[Y])^2 \\ &\overset{\text{Lem}}{\underset{3.7b)}{=}} V[X] + V[Y] + 2(E[XY] - E[X]E[Y]) \\ &\overset{a)}{=} V[X] + V[Y] + 2Cov(X, Y). \end{aligned}$$

■

BEMERKUNG: Für $Y := X$ folgt aus der Behauptung:

$$V[2X] = V[X + X] = V[X] + V[X] + 2Cov(X, X) = 2V[X] + 2V[X] = 4V[X],$$

was auch nach Lemma 3.7 a) korrekt ist.

Aufgabe 57 [Fangfrage zu Stichproben-Korrelation und (lineare) Gleichläufigkeit]
In der Vorlesung haben wir in Lemma 3.10 b) gezeigt, dass wir eine lineare Gleichläufigkeit zwischen X und Y haben, wenn der Absolutbetrag der Korrelation gleich 1 ist: $|Cor(X, Y)| = 1$. Im Fall von "+1" liegen bei zentrierten Zufallsgrössen (fast) alle Punkte auf einer Geraden durch den Nullpunkt mit positiver Steigung, im Fall von -1 liegen (fast) alle Punkte auf einer Geraden durch den Nullpunkt mit negativer Steigung.

Falls nun alle Punkte auf der x-Achse liegen, bedeutet dass: $Y(\omega) = 0$, für alle $\omega \in \Omega$. Es folgt damit sofort: $V[Y] = 0$, also ist in diesem Fall $Cor(X, Y)$ gar nicht definiert, da $Cor(X, Y)$ nur für $V[X], V[Y] > 0$ definiert ist.

Aufgabe 58 [R, Korrelationskoeffizient "von Hand" berechnen]

a)

```
> PN <- 3
> u <- rexp(1000,PN)
> v <- log(u)
> cor(u,v)
[1] 0.7929965
```

Der Zusammenhang zwischen `u` und `v` ist nicht linear, also ist nach Lemma 3.10 b) zu erwarten, dass $|$`cor(u,v)`$| \neq 1$, was auch zutrifft. Da weiter log immer eine positive Steigung hat, ist eine positive Korrelation zu erwarten. Auch dies wird durch die obige R-Ausgabe bestätigt.

Weiter wenn die Daten `u` nahe bei einander liegen würden und log dort etwa linear ist (zum Beispiel lokal um 1), dann wäre `cor(u,v)` nahe von 1. Dies aufgrund von Lemma 3.10 b).

b)

```
> #Wir definieren die Stichproben-Varianz
> #s2 wie in Aufgabe 52:
> s2 <- function(x) return(1/length(x) * sum((x-mean(x))^2))
>
> #Nun definieren wir unsere eigenen cov-
> #und cor-Funktionen: mycov und mycor:
> #Wir benutzen dazu Aufgabe 56 a)
> mycov <- function(x,y) return(mean(x*y) - mean(x)*mean(y))
> mycor <- function(x,y) return(mycov(x,y)/sqrt(s2(x)*s2(y)))
>
> mycor(u,v)
[1] 0.7929965
```

Wir erhalten mit unserer eigenen Funktion `mycor` dasselbe Resultat wie in a). Wenn man die Funktion `var` statt `s2` verwendet, teilt R bei der Stichproben-Varianz bekanntlich durch "$n-1$" statt "n", was natürlich auch hier zu einem etwas anderen Resultat führt. Damit Lemma 3.10 b) auch mit der Stichproben-Covarianz gilt, muss aber `s2` verwendet werden.

Aufgabe 59 [R, Korrelationskoeffizient und lineare Gleichläufigkeit]

a)

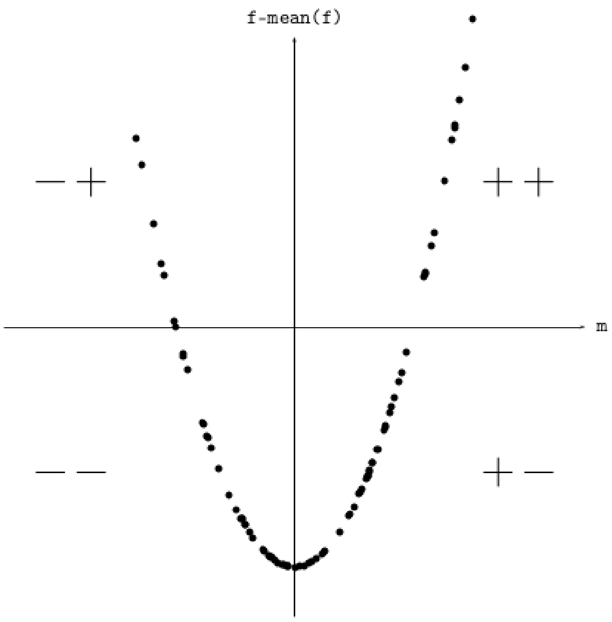
```
> m <- rnorm(100000,0,sqrt(4))
> n <- 5+ 2*m
> cor(m,n)
[1] 1
```

Der Zusammenhang zwischen m und n ist linear, also muss nach Lemma 3.10 b) gelten: $|Cor(m,n)| = 1$. Da die Steigung positiv ist, folgt sogar $Cor(X,Y) = 1$. Dies begründet natürlich obiges Resultat.

b)

```
> f=m^2
> cor(m, f)
[1] 0.002856787
>
```

Die Korrelation ist also fast Null. Eine kurze Erklärung liefert die folgende Skizze (mit einer 100-er Stichprobe für eine bessere Übersicht):



Die Werte heben sich also in etwa gegenseitig auf! Wir vermuten also eine Korrelation nahe bei Null.

Eine exaktere Begründung liefert dann die nächste Teilaufgabe.

c) Sei M eine $\mathcal{N}(0,4)$-verteilte Zufallsgrösse und $F := M^2$. Nun gilt:

$$Cov(M,F) \overset{56a)}{=} E[MF] - \underbrace{E[M]}_{=0} E[F] = E[M^3] = \frac{1}{\sqrt{2\pi}\cdot 2}\int_{-\infty}^{\infty} x^3 e^{-\frac{x^2}{2\cdot 2^2}}\,dx \underset{\text{metrie}}{\overset{\text{Sym-}}{=}} 0.$$

Wir haben dabei benutzt, dass der Integrand eine ungerade Funktion ist.

Wir schliessen: M und F sind unkorreliert. Dies erklärt das Resultat aus b).