Vorlesung 9b

Bedingte Varianz

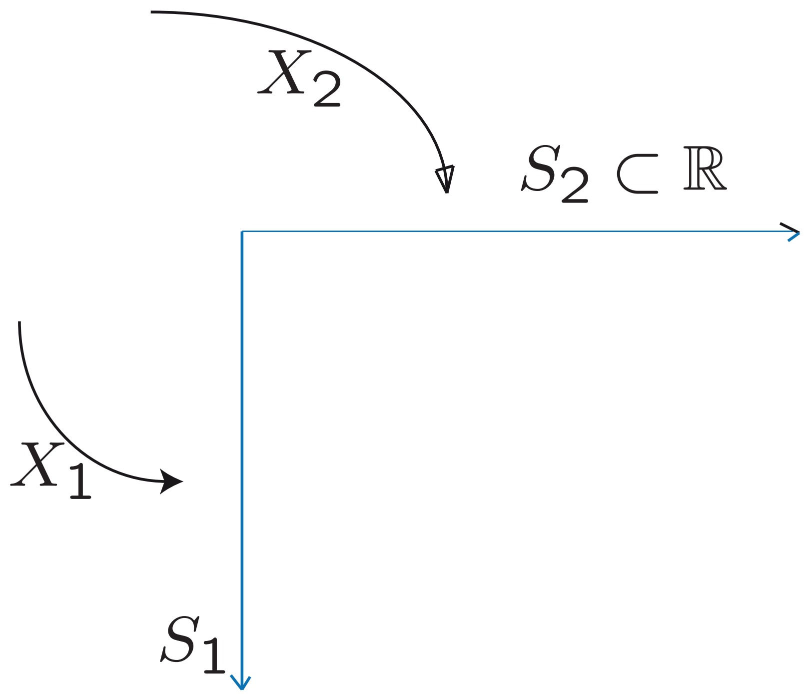
und

Addieren von unabhängigen Zufallsvariablen

- zweistufig aufgefasst

1. Definition der bedingten Varianz

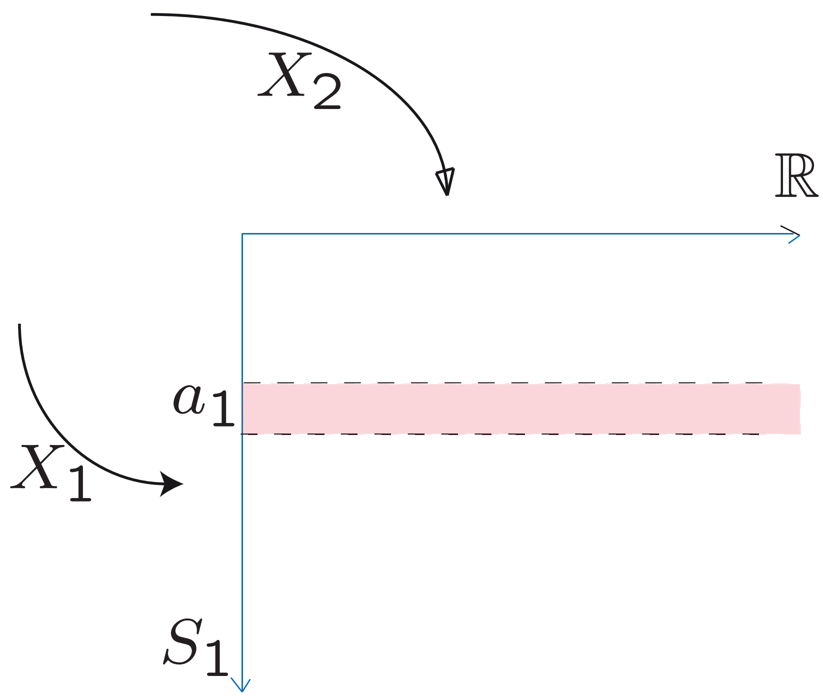
(Buch S. 90)



(X_1, X_2) sei (hier der Einfachheit halber) diskret:

S_1 und S_2 endlich oder abzählbar

X_2 reellweertig mit $\mathbf{P}(X_2 \in S_2) = 1$



Die auf das Ereignis $\{X_1 = a_1\}$ bedingte Varianz

wollen wir verstehen als

die Varianz innerhalb der Zeile namens a_1

Wir definieren die

bedingte Varianz von X_2, gegeben $\{X_1 = a_1\}$ als

$$\mathbf{Var}_{a_1}[X_2] := \mathbf{E}_{a_1}\Big[(X_2 - \mathbf{E}_{a_1}[X_2])^2\Big]$$

Dies ist also die Varianz der Wahrscheinlichkeitsverteilung mit den Gewichten $P(a_1, a_2)$, $a_2 \in S_2$ $(\subset \mathbb{R})$.

Dabei ist $P(a_1, .)$ die Übergangsverteilung “in Zeile a_1 ”, siehe Vorlesung 8b.

Die bedingte Varianz ist somit

der bedingte Erwartungswert der

quadratischen Abweichung vom bedingten Erwartungswert.*

Wie sieht es aus, wenn man hier

den bedingten Erwartungswert $\mathbf{E}_{a_1}[X_2]$

durch einen anderen "Prognosewert" $h(a_1)$ ersetzt?

*Den bedingten Erwartungswert $\mathbf{E}_{a_1}[X_2]$ kann man auffassen als einen *Prognosewert* für X_2, gegeben $\{X_1 = a_1\}$.

2. Bedingter Erwartungswert einer quadratischen Abweichung

(Buch S. 90)

Wir erinnern uns an die “hilfreiche Formel für die Varianz”:

$$\mathbf{E}[Y^2] = \mathbf{Var}[Y] + (\mathbf{E}[Y])^2.$$

Verschieben um eine Konstante ändert die Varianz nicht:

$$\mathbf{E}[(Y - c)^2] = \mathbf{Var}[Y] + (\mathbf{E}[Y] - c)^2.$$

Was der Varianz recht ist, ist der bedingten Varianz billig:

Für alle $a_1 \in S_1$ und beliebiges $h(a_1) \in \mathbb{R}$ ist

$$\mathbf{E}_{a_1}[(X_2 - h(a_1))^2] = \mathbf{Var}_{a_1}[X_2] + (\mathbf{E}_{a_1}[X_2] - h(a_1))^2.$$

$$\mathbf{E}_{a_1}[(X_2 - h(a_1))^2] = \mathbf{Var}_{a_1}[X_2] + (\mathbf{E}_{a_1}[X_2] - h(a_1))^2.$$

Anstelle von a_1 können wir die Zufallsvariable X_1 einsetzen:

$$\mathbf{E}_{X_1}[(X_2 - h(X_1))^2] = \mathbf{Var}_{X_1}[X_2] + (\mathbf{E}_{X_1}[X_2] - h(X_1))^2.$$

Bilde in dieser Gleichheit den Erwartungswert

und verwende links dessen Zerlegung nach X_1.

(also die Formel auf V9a Folie 14, hier mit

$g(a_1, a_2) := (a_2 - h(a_1))^2$:

$$(*) \qquad \mathbf{E}[(X_2 - h(X_1))^2]$$
$$= \mathbf{E}[\mathbf{Var}_{X_1}[X_2]] + \mathbf{E}[(\mathbf{E}_{X_1}[X_2] - h(X_1))^2].$$

$$(*) \qquad \mathbf{E}[(X_2 - h(X_1))^2]$$

$$= \mathbf{E}[\mathbf{Var}_{X_1}[X_2]] + \mathbf{E}[(\mathbf{E}_{X_1}[X_2] - h(X_1))^2].$$

Diese Formel können wir so lesen:

Der erwartete quadratische Prognosefehler
bei Verwendung der Prognose $h(X_1)$ ist die Summe
aus dem Erwartungswert der bedingten Varianz
und dem "erwarteten quadratischen Bias",
der dadurch entsteht, dass man $h(X_1)$ anstelle von
$\mathbf{E}_{X_1}[X_2]$ als Prognose von X_2 verwendet.

3. Die bedingte Erwartung als beste Prognose im quadratischen Mittel

(Buch S. 90)

Satz:

Sei X_2 reellwertige Zufallsvariable mit $\mathbf{E}[X_2^2] < \infty$.

Dann minimiert die bedingte Erwartung $\mathbf{E}_{X_1}[X_2]$
unter allen reellwertigen Zufallsvariablen der Form $h(X_1)$
den erwarteten quadratischen Abstand

$$\mathbf{E}\big[(X_2 - h(X_1))^2\big].$$

Beweis. Am Ende von Abschnitt 2 hatten wir gesehen:

$$(*) \qquad \mathbf{E}[(X_2 - h(X_1))^2]$$

$$= \mathbf{E}[\mathbf{Var}_{X_1}[X_2]] + \mathbf{E}[(\mathbf{E}_{X_1}[X_2] - h(X_1))^2].$$

Der Satz über die Positivität des Erwartungswerts impliziert:

Der letzte Term rechts wird minimal (nämlich 0)

genau dann, wenn

$$\mathbf{P}\Big(h(X_1) = \mathbf{E}_{X_1}[X_2]\Big) = 1.$$

Äquivalent dazu:

$h(a_1) = \mathbf{E}_{a_1}[X_2]$ für alle a_1 mit $\mathbf{P}(X_1 = a_1) > 0$. $\square$

Fazit:

1. Unter allen Zahlen $h(a_1)$
ist der bedingte Erwartungswert $\mathbf{E}_{a_1}[X_2]$
diejenige Zahl, für die $\mathbf{E}_{a_1}[(X_2 - h(a_1))^2]$ minimal wird.

2. Unter allen Zufallsvariablen der Form $h(X_1)$
ist die bedingte Erwartung $\mathbf{E}_{X_1}[X_2]$
diejenige, für die

$$\mathbf{E}[\mathbf{E}_{X_1}[(X_2 - h(X_1))^2]] = \mathbf{E}[(X_2 - h(X_1))^2]$$

minimal wird.

4. Zerlegung der Varianz

(Buch S. 90)

Wieder verwenden wir die Formel $(*)$

vom Ende von Abschnitt 2:

$$(*) \qquad \mathbf{E}[(X_2 - h(X_1))^2]$$

$$= \mathbf{E}[\mathbf{Var}_{X_1}[X_2]] + \mathbf{E}[(\mathbf{E}_{X_1}[X_2] - h(X_1))^2].$$

Jetzt setzen wir

$$h(a_1) := \mathbf{E}[X_2], \quad a_1 \in S_1.$$

Dann wird $(*)$ zu

$$\mathbf{E}[(X_2 - \mathbf{E}[X_2])^2]$$
$$= \mathbf{E}[\mathbf{Var}_{X_1}[X_2]] + \mathbf{E}[(\mathbf{E}_{X_1}[X_2] - \mathbf{E}[X_2])^2].$$

Wegen

$$\mathbf{E}[X_2] = \mathbf{E}[\mathbf{E}_{X_1}[X_2]]$$

wird der zweite Term rechts zu

$$\mathbf{Var}\Big[\mathbf{E}_{X_1}[X_2].\Big]$$

Insgesamt wird damit die Formel $(*)$ zu

$$\boxed{\mathbf{Var}[X_2] = \mathbf{E}\Big[\mathbf{Var}_{X_1}[X_2]\Big] + \mathbf{Var}\Big[\mathbf{E}_{X_1}[X_2]\Big]}$$

Zerlegung der Varianz von X_2 nach X_1.

$$\mathbf{Var}[X_2] = \mathbf{E}\Big[\mathbf{Var}_{X_1}[X_2]\Big] + \mathbf{Var}\Big[\mathbf{E}_{X_1}[X_2]\Big]$$

Zum Merken:

Die Varianz von X_2

ist die Summe aus dem

Erwartungswert der bedingten Varianzen

und der Varianz der bedingten Erwartungswerte

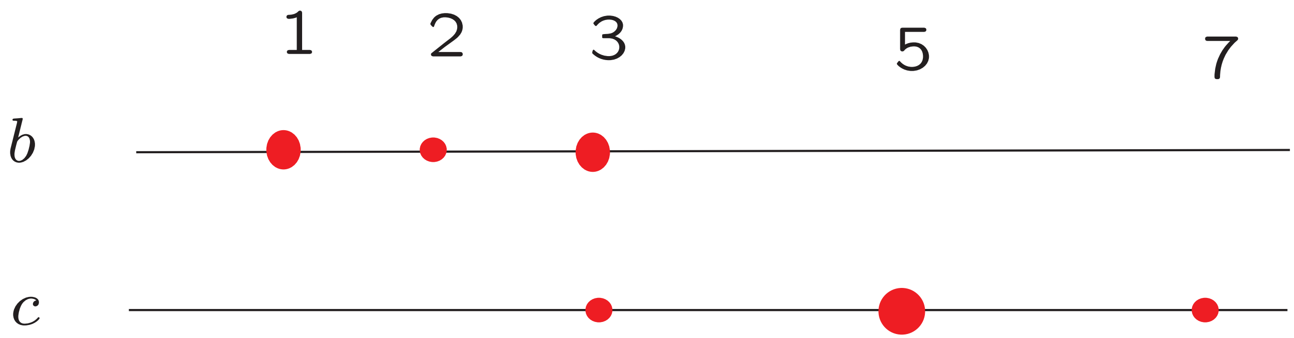
(Variabilität innerhalb der Zeilen

plus Variabilität zwischen den Zeilen).

Wir illustrieren dies mit einem kleinen Besipiel:

Die Übergangsmatrix P sei

	1	2	3	5	7
b	0.4	0.2	0.4	0	0
c	0	0	0.2	0.6	0.2



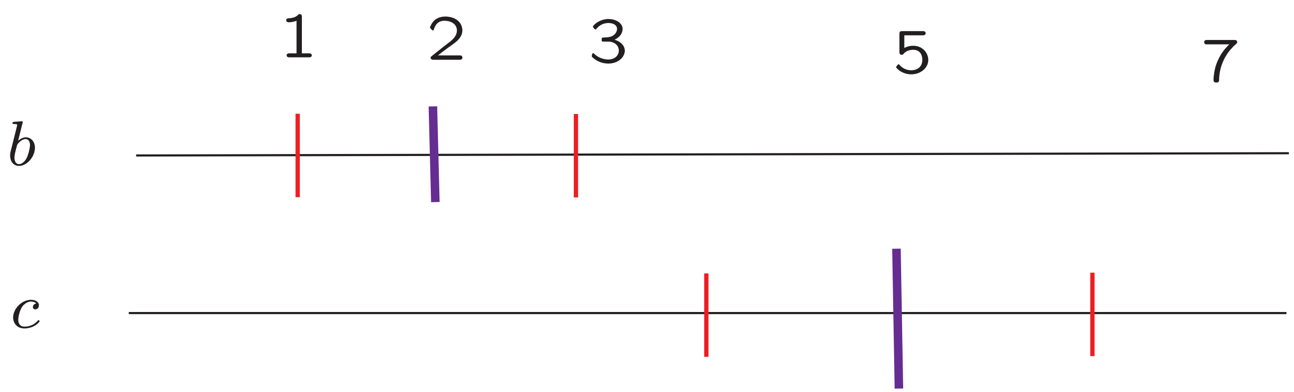
Dann gilt:

$\mathbf{E}_b[X_2] = 2, \quad \mathbf{E}_c[X_2] = 5,$

$\mathbf{Var}_b[X_2] = 0.8 \cdot 1^2 = 0.8,\ \mathbf{Var}_c[X_2] = 0.4 \cdot 2^2 = 1.6.$

$\mathbf{E}_b[X_2] = 2$, $\mathbf{E}_c[X_2] = 5$,

$\mathbf{Var}_b[X_2] = 0.8$, $\mathbf{Var}_c[X_2] = 1.6$.



Die Startgewichte seien $\rho(b) = 0.3$, $\rho(c) = 0.7$. Damit:

- Erwartungswert der bedingten Varianzen: $0.3 \cdot 0.8 + 0.7 \cdot 1.6$

- Varianz der bedingten Erwartung: $0.3 \cdot 0.7 \cdot (5 - 2)^2$

Deren Summe ist $\mathbf{Var}[X_2] = 3.25$.

Wir erklären noch das Ergebnis

$$\mathbf{Var}[e(X_1)] = 0.3 \cdot 0.7 \cdot (5-2)^2$$

für $e(X_1) := \mathbf{E}_{X_1}[X_2]$

aus der vorigen Folie.

$e(X_1)$ ist hier eine binäre Zufallsvariable

und hat dieselbe Verteilung wie $2 + Z(5-2)$,

wobei Z ein Münzwurf mit $p = 0.7$ ist ...

Beispiel: Summe aus einer zufälligen Anzahl unabhängiger Summanden.

$$Y := \sum_{i=1}^{N} Z_i$$

mit $Z_1, Z_2, \ldots$ unabhängig, identisch verteilt und unabhängig von N.

$$\mu := \mathbf{E}[Z_1],\ \sigma^2 := \mathbf{Var}[Z_1]$$

Aufgabe: Berechne $\mathbf{E}[Y]$ und $\mathbf{Var}[Y]$ aus $\mathbf{E}[N]$, $\mathbf{Var}[N]$, μ und σ^2.

$$Y = \sum_{i=1}^{N} Z_i, \quad \mu := \mathbf{E}[Z_1], \ \sigma^2 := \mathbf{Var}[Z_1].$$

Wir nehmen N als erste und Y als zweite Stufe:

$$\mathbf{E}_n[Y] = n\mu \ , \quad \mathbf{Var}_n[Y] = n\sigma^2 \ .$$

$$\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}_N[Y]] = \mathbf{E}[N\mu] = \mathbf{E}[N] \cdot \mu.$$

$$\mathbf{Var}[Y] = \mathbf{E}\big[\mathbf{Var}_N[Y]\big] + \mathbf{Var}\big[\mathbf{E}_N[Y]\big]$$
$$= \mathbf{E}[N] \cdot \sigma^2 + \mathbf{Var}[N] \cdot \mu^2.$$

5. Addieren von unabhängigen ZV'en
– zweistufig aufgefasst:

Der diskrete Fall

(Buch S. 92)

Y und Z seien unabhängige $\mathbb{Z}$-wertige Zufallsvariable.

Wie ist $Y + Z$ verteit?

Wir können $Y + Z$ auffassen als

zweite Stufe eines Zufallsexperiments.

Die erste Stufe ist Y.

Gegeben $\{Y = a\}$ ist

$$X_2 := Y + Z$$

so verteilt wie $a + Z$.

$$\mathbf{P}(Y = a, Y + Z = b) = \mathbf{P}(Y = a,\ a + Z = b)$$
$$= \mathbf{P}(Y = a)\,\mathbf{P}(a + Z = b)$$

Wir sehen hier die Produktformel wieder!

Summation über a ergibt die “totale Wahrscheinlichkeit”:

$$\mathbf{P}(Y + Z = b) = \sum_a \mathbf{P}(Y = a)\,\mathbf{P}(Z = b - a)$$

(Zerlegung von $\mathbf{P}(Y + Z = b)$ nach den Ausgängen von Y,

“Zerlegung nach dem ersten Schritt”)

$$\mathbf{P}(Y+Z=b)=\sum_a \mathbf{P}(Y=a)\,\mathbf{P}(Z=b-a)$$

Beispiel:

Y, Z unabhängig und Geom(p)-verteilt

$$\mathbf{P}(Y+Z=b)=\sum_{a=1}^{b-1} pq^{a-1}pq^{b-a-1}$$

$$=(b-1)p^2q^{b-2}, \quad b=2,3,\dots$$

Dies sind die Gewichte der sogenannten

negativen Binomialverteilung mit Parametern 2, p

Diese ist die Verteilung der Anzahl der Versuche

in einem p-Münzwurf bis einschließlich zum zweiten Erfolg.

6. Addieren von unabhängigen ZV'en
– zweistufig aufgefasst:

Der Fall mit Dichten

(Buch S. 92)

Im diskreten Fall hatten wir

$$\mathbf{P}(Y = a, Y + Z = b) = \mathbf{P}(Y = a, a + Z = b)$$

$$= \mathbf{P}(Y = a)\,\mathbf{P}(a + Z = b)$$

$$= \mathbf{P}(Y = a)\,\mathbf{P}(Z = b - a)$$

Haben Y und Z die Dichten $f(y)\,dy$ und $g(z)\,dz$,

so bekommt man analog

die gemeinsame Dichte von $(Y, Y + Z)$:

$$\mathbf{P}(Y \in da, Y + Z \in db) = \mathbf{P}(Y \in da, a + Z \in db)$$

$$= \mathbf{P}(Y \in da)\,\mathbf{P}(a + Z \in db)$$

$$= f(a)da\;\, g(b - a)\,db$$

$$\boxed{\mathbf{P}(Y \in da, Y + Z \in db) = f(a)\; g(b-a)\; da\, db}$$

Integration über a gibt die Dichte von $Y + Z$:

$$\mathbf{P}(Y + Z \in db) = \Big(\int f(a)\, g(b-a)\; da\Big)\, db\;.$$

Beispiel: Für Y und Z unabhängig und Exp(1)-verteilt ist

$$\mathbf{P}(Y + Z \in db) = \Big(\int_0^b e^{-a} e^{-(b-a)}\; da\Big)\, db$$
$$= be^{-b}\, db, \quad b \ge 0.$$

(Dichte der Gamma(2)-Verteilung)

7. Noch ein Beispiel für die Zerlegung der Varianz:

Die Varianz einer Summe

von zwei unabhängigen Zufallsvariablen

Y und Z seien unabhängige, reellwertige Zufallsvariable mit endlicher Varianz. Wir interpretieren die uns schon bekannte Formel $\sigma^2_{Y+Z} = \sigma^2_Y + \sigma^2_Z$ jetzt noch einmal, als Zerlegung der Varianz der zweiten Stufe nach der ersten:

Die erste Stufe ist $X_1 := Y$, die zweite ist $X_2 := Y + Z$. Die Übergangsverteilung $P(a_1, .)$ ist hier also die die Verteilung von $a_1 + Z$. (vgl. Abschnitte 5 und 6).

Daraus folgt:

Die bedingte Erwartung von $Y + Z$ gegeben Y ist

$$Y + \mu_Z.$$

Die Varianz der bedingten Erwartung ist also σ_Y^2.

Die bedingte Varianz von $Y + Z$, gegeben Y, ist

$$\sigma_Z^2.$$

Der Erwartungswert der bedingten Varianz ist also σ_Z^2.

Mit Blick auf das Resultat in Abschnitt 4 ergibt sich:

$$\sigma_{Y+Z}^2 = \sigma_Z^2 + \sigma_Y^2.$$