Vorlesung 9b

Bedingte Erwartung und bedingte Varianz

1. Zerlegung eines Erwartungswertes nach der ersten Stufe

Wie in der vorigen Vorlesung betrachten wir die

gemeinsame Verteilung von zwei Zufallsvariablen X_1, X_2,

aufgebaut aus der Verteilung von X_1

und den Übergangsverteilungen:

$$\mathbf{P}(X_1 = a_1, X_2 = a_2) = \mathbf{P}(X_1 = a_1)\,\mathbf{P}_{a_1}(X_2 = a_2)$$

$$\mu(a_1, a_2) = \rho(a_1) P(a_1, a_2)$$

Auch den Erwartungswert

einer reellwertigen Zufallsvariablen $g(X_1, X_2)$

kann nach der ersten Stufe zerlegt werden.

Sei $g : S_1 \times S_2 \to \mathbb{R}$.

Wir betrachten die Zufallsvariable $g(X_1, X_2)$.

Für $a_1 \in S_1$ setzen wir

$$\mathbf{E}_{a_1}\big[g(X_1, X_2)\big] := \sum_{a_2 \in S_2} g(a_1, a_2)\mathbf{P}_{a_1}(X_2 = a_2)$$

und nennen diese Zahl den

bedingten Erwartungswert von $g(X_1, X_2)$,

gegeben $\{X_1 = a_1\}$.

Merke:

Der bedingte Erwartungswert

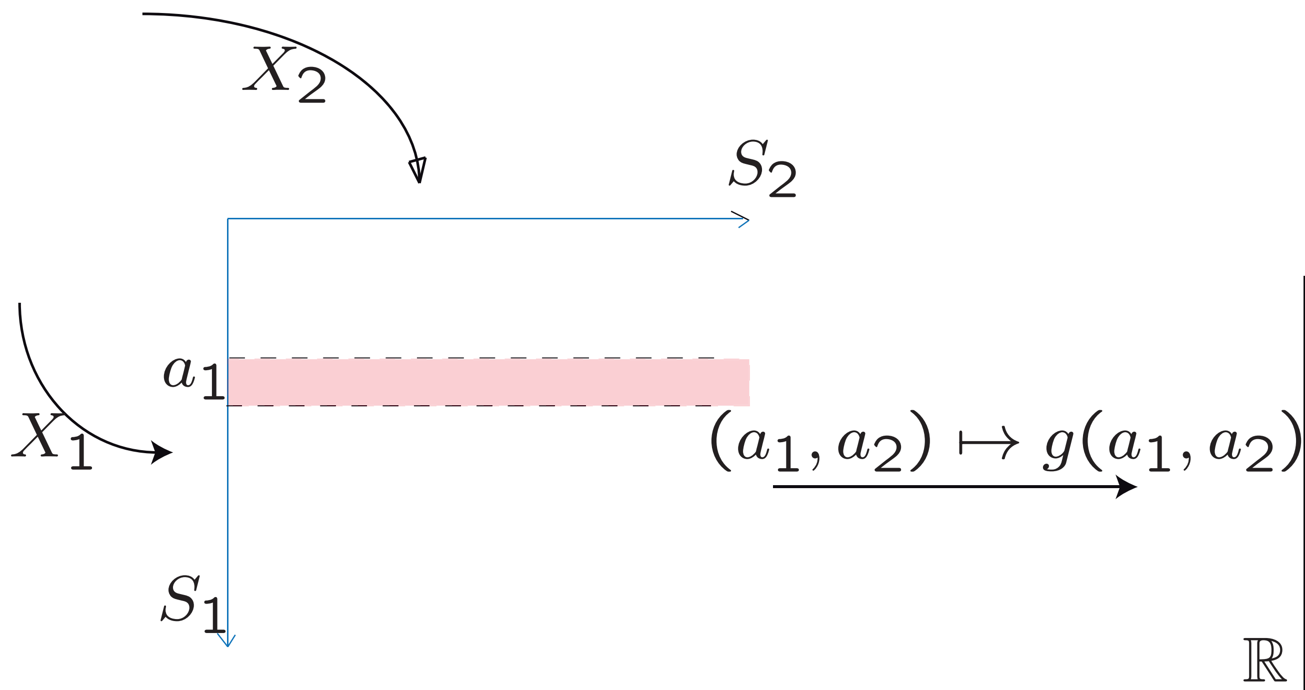
$$\mathbf{E}_{a_1}\big[g(X_1, X_2)\big]$$

wird gebildet mit der Übergangsverteilung $P(a_1, .)$,

also mit den Wahrscheinlichkeitsgewichten,

die die Zeile $P(a_1, .)$ der Matrix P bilden:

$$\mathbf{E}_{a_1}\big[g(X_1, X_2)\big] = \sum_{a_2 \in S_2} g(a_1, a_2) P(a_1, a_2)$$


X_2
S_2
a_1
X_1
$(a_1, a_2) \mapsto g(a_1, a_2)$
S_1
$\mathbb{R}$

Ähnlich wie die Verteilungsgewichte von (X_1, X_2) lässt sich auch der Erwartungswert $\mathbf{E}[g(X_1, X_2)]$ nach den Ausgängen von X_1 zerlegen.

(Zerlegung des Erwartungswerts nach der ersten Stufe)

$$\mathbf{E}\big[g(X_1, X_2)\big]$$

$$= \sum_{a_1 \in S_1} \sum_{a_2 \in S_2} g(a_1, a_2)\, \mathbf{P}(X_1 = a_1, X_2 = a_2)$$

$$= \sum_{a_1 \in S_1} \sum_{a_2 \in S_2} g(a_1, a_2)\, \mathbf{P}(X_1 = a_1)\, \mathbf{P}_{a_1}(X_2 = a_2)$$

$$= \sum_{a_1 \in S_1} \mathbf{P}(X_1 = a_1) \mathbf{E}_{a_1}\big[g(X_1, X_2)\big]$$

$$= \sum_{a_1 \in S_1} \mathbf{P}(X_1 = a_1) \mathbf{E}_{a_1}\big[g(a_1, X_2)\big]$$

$$= \mathbf{E}\big[\mathbf{E}_{X_1}[g(X_1, X_2)]\big].$$

“Zerlegung des Erwartungswertes nach der ersten Stufe”

Merke: Der bedingte Erwartungswert

$$\mathbf{E}_{a_1}\left[g(X_1, X_2)\right]$$

ist eine Zahl.

$$\mathbf{E}_{X_1}\left[g(X_1, X_2)\right]$$

ist eine Zufallsvariable.

Wir nennen diese Zufallsvariable die

bedingte Erwartung von $g(X_1, X_2)$ gegeben X_1.

$$\mathbf{E}\big[g(X_1, X_2)\big] = \mathbf{E}\big[\mathbf{E}_{X_1}[g(X_1, X_2)]\big]$$

Ist X_2 reellwertig (also $S_2 \subset \mathbb{R}$),

dann ergibt sich als Spezialfall (mit $g(a_1, a_2) := a_2$)

$$\mathbf{E}_{a_1}[X_2] = \sum_{a_2 \in S_2} a_2 \, \mathbf{P}_{a_1}(X_2 = a_2) \, ,$$

Diese Zahl nennen wir den

Erwartungswert von X_2, gegeben $X_1 = a_1$.

Wir haben dann die einprägsame Formel

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

(Zerlegung des Erwartungswertes von X_2 nach X_1.)

2. Ein Beispiel: Suchen in Listen.

n Namen werden in r Listen einsortiert. Dadurch ergibt sich

ein r-Tupel $k = (k_1, \ldots, k_r)$ von Listenlängen

(eine "Besetzung" k der Plätze $1, \ldots, r$)

Jeder Name steht in seiner Liste Nr. j

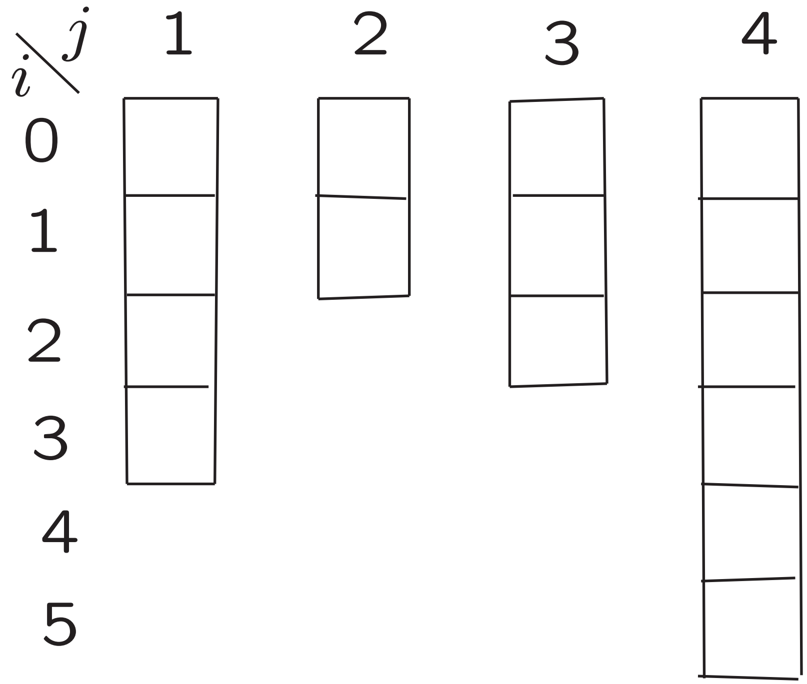
an einer der Stellen $i = 0, \ldots, k_j - 1$.

Vorstellung: Die Listennummer entspricht dem

Anfangsbuchstaben des Namens.

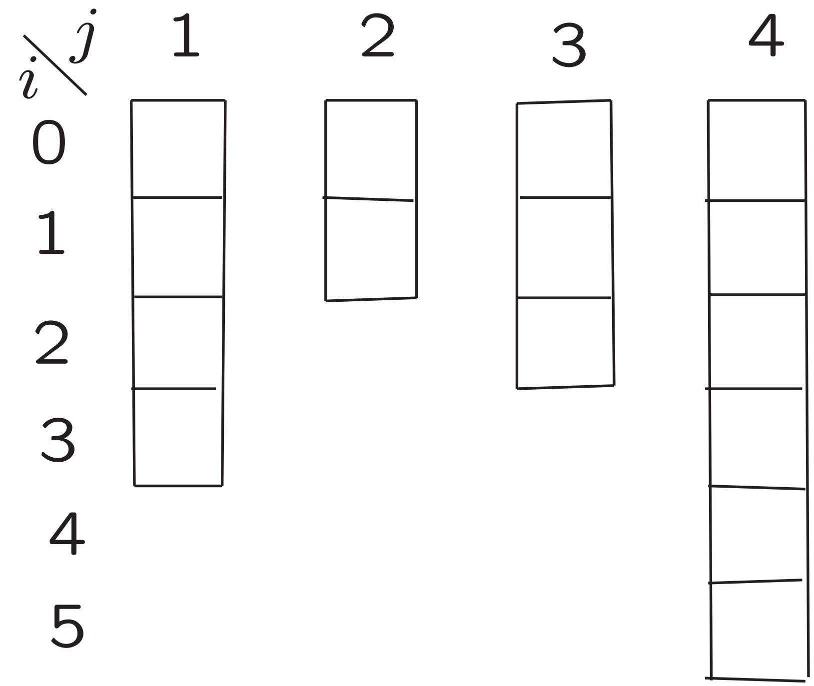
Für ein Alphabet mit $r = 4$ Buchstaben

(und bei $n = 15$ Namen) ist *eine* mögliche Besetzung:



$$n = 15, r = 4$$

$$k = (k_1, k_2, k_3, k_4) = (4, 2, 3, 6)$$

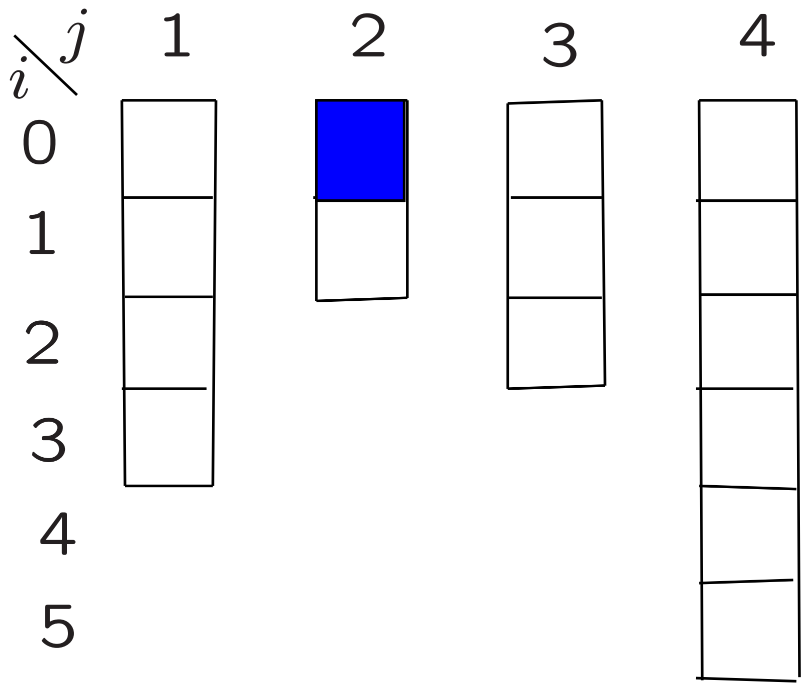


Erste Frage: Was ist für gegebene Listenlängen (k_j)

der Erwartungswert der Stellennummer M

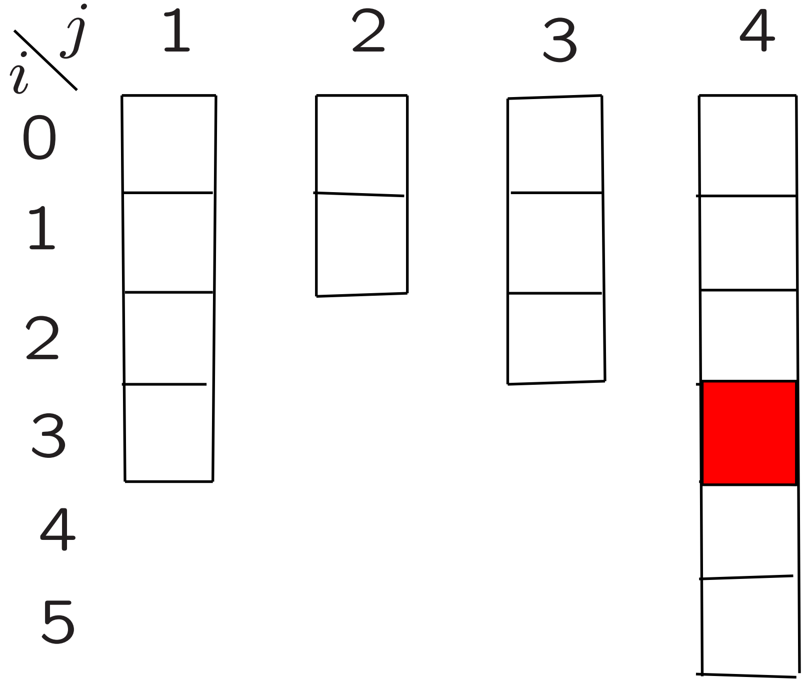
(der “Suchtiefe”)

eines rein zufällig aus den n herausgegriffenen Names?



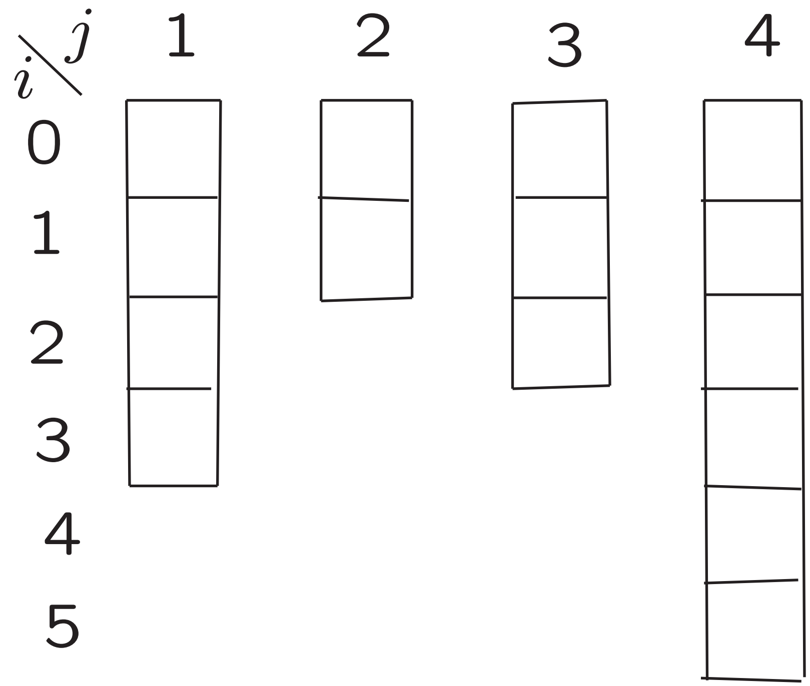
Liste $j = 2$,

Tiefe $i = 0$.



Liste $j = 4$,

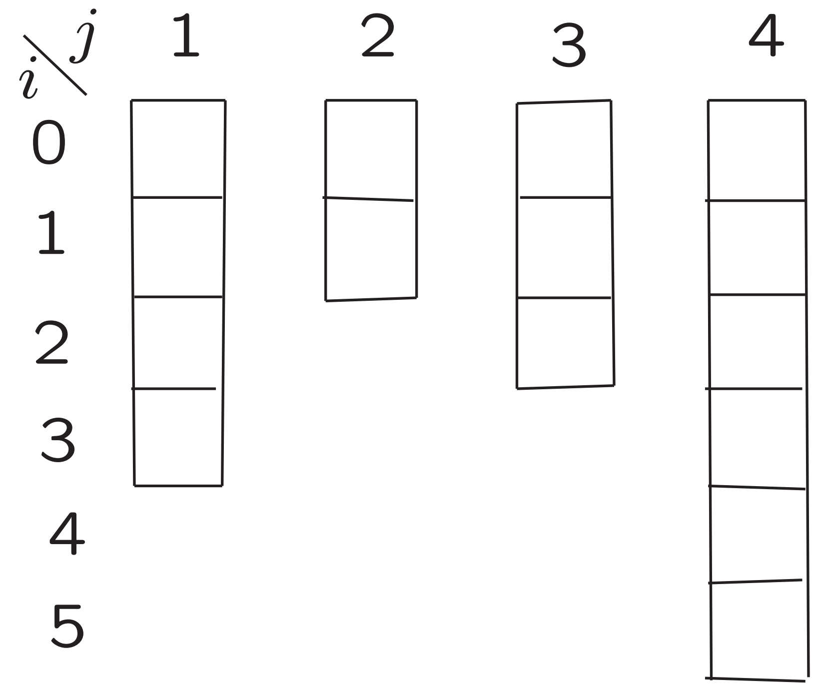
Tiefe $i = 3$.



Was ist *bei gegebenem* k

der Erwartungswert der Suchtiefe M

eines rein zufällig aus den n herausgegriffenen Names?



Die Antwort ist

$$\mathbf{E}_k[M] = \frac{1}{n} \sum_{j=1}^{r} \sum_{i=0}^{k_j-1} i = \frac{1}{n} \sum_{j=1}^{r} \frac{k_j(k_j-1)}{2} \; .$$

Wir betrachten jetzt ein

stochastisches Modell für die erste Stufe:

Annahme:

Die zufällige Besetzung $Z = (Z_1, \ldots, Z_r)$

kommt durch n-maliges Würfeln

mit den Gewichten $p_1, \ldots, p_r$ zustande.

Z ist multinomial (n, $p_1, \ldots, p_r$)-verteilt.

(Vorstellung: Die n Namen

sind eine Stichprobe aus einer großen Population

mit bekannter Verteilung der Anfangsbuchstaben.)

Aus den n Namen wird rein zufällig einer herausgegriffen.

Es sei M die Stelle, die er in seiner Liste einnimmt.

Aufgabe: Berechne $\mathbf{E}[M]$.

(Dieser Erwartungswert beschreibt die mittlere Suchzeit

(Suchtiefe)

für einen aus den Listen zufällig gewählten Namen.)

Der Erwartungswert von M, gegeben $Z = k$, war

$$\mathbf{E}_k[M] = \frac{1}{n} \sum_{j=1}^{r} \frac{k_j(k_j - 1)}{2} .$$

Mit der oben hergeleitetet Zerlegung des Erwartungswertes

$$\mathbf{E}[M] = \mathbf{E}\big[\mathbf{E}_Z[M]\big]$$

erhalten wir

$$\mathbf{E}[M] = \frac{1}{n} \sum_{j=1}^{r} \mathbf{E}\Big[\frac{Z_j(Z_j - 1)}{2}\Big] .$$

$$\mathbf{E}[M] = \frac{1}{n} \sum_{j=1}^{r} \mathbf{E}\Big[\frac{Z_j(Z_j - 1)}{2}\Big]$$

Nach Annahme ist Z_j Binomial(n, p_j)-verteilt.

Mit der Formel $\mathbf{Var}[Z] = \mathbf{E}[Z^2] - (\mathbf{E}[Z])^2$ folgt

$$\mathbf{E}[Z_j(Z_j - 1)] = \mathbf{Var}[Z_j] + \mathbf{E}[Z_j]^2 - \mathbf{E}[Z_j]$$

$$= np_j(1 - p_j) + (np_j)^2 - np_j = p_j^2\, n(n-1),$$

$$\mathbf{E}[M] = \frac{n-1}{2}(p_1^2 + \cdots + p_r^2)\,.$$

$$\mathbf{E}[M] = \frac{n-1}{2}(p_1^2 + \cdots + p_r^2) \ .$$

Im Fall uniformer Gewichte

$$p_1 = \cdots = p_r = 1/r$$

ergibt sich

$$\mathbf{E}[M] = \frac{n-1}{2r} \ .$$

Eine Modifikation des vorigen Beispiels:

Sei Z wieder multinomial $(n,\, p_1, \ldots, p_r)$-verteilt,

J sei unabhängig von Z, mit $\mathbf{P}(J = j) = p_j, \quad j = 1, \ldots, r$.

Berechne den Erwartungswert von $X := Z_J$.

(Dieser Erwartungswert beschreibt die mittlere Suchzeit

nach einem in den Listen nicht vorhandenen Namen)

Wir zerlegen $\mathbf{E}[X]$ nach den Ausgängen von Z:

$$\mathbf{E}[X] = \mathbf{E}[\mathbf{E}_Z[X]] = \sum_k \mathbf{P}(Z = k)\, \mathbf{E}_k[X]$$

$$= \sum_k \mathbf{P}(Z = k) \sum_{j=1}^{r} p_j k_j$$

$$= \sum_{j=1}^{r} p_j \sum_k \mathbf{P}(Z = k) k_j$$

$$= \sum_{j=1}^{r} p_j\, \mathbf{E} Z_j = \sum_{j=1}^{r} p_j\, n p_j = n \sum_{j=1}^{r} p_j^2 .$$

Im Fall uniformer Gewichte

$$p_1 = \cdots = p_r = 1/r$$

ergibt sich

$$\mathbf{E}[X] = \frac{n}{r}.$$

Im Vergleich dazu war (siehe voriges Besipiel)

die mittlere Suchtiefe eines rein zufällig aus den n

herausgegriffenen Namens

$$\mathbf{E}[M] = \frac{n-1}{2r}.$$

3. Bedingte Varianz

Für reellwertiges X_2 definieren wir die

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

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

Dies is die Varianz der Wahrscheinlichkeitsverteilung

mit den Gewichten $P(a_1, a_2)$, $a_2 \in S_2$.

Auch die Varianz $\mathbf{Var}[X_2]$ lässt sich, wie wir gleich einsehen werden, nach der ersten Stufe zerlegen, und zwar so:

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

Dies ist die **Formel von der Zerlegung der Varianz**.

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

Zum Merken:

Die Varianz von X_2

ist die Summe aus dem

Erwartungswert der bedingten Varianzen

und der Varianz der bedingten Erwartungswerte

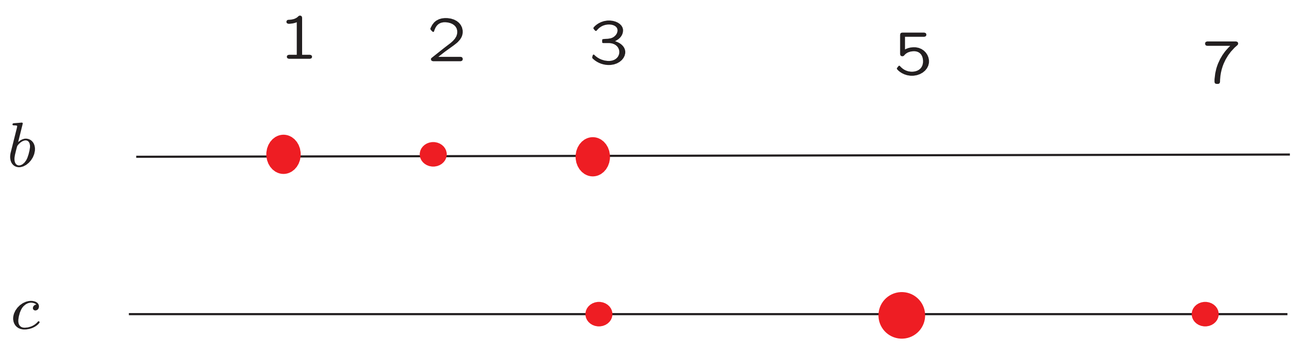
(Variabilität innerhalb der Zeilen

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



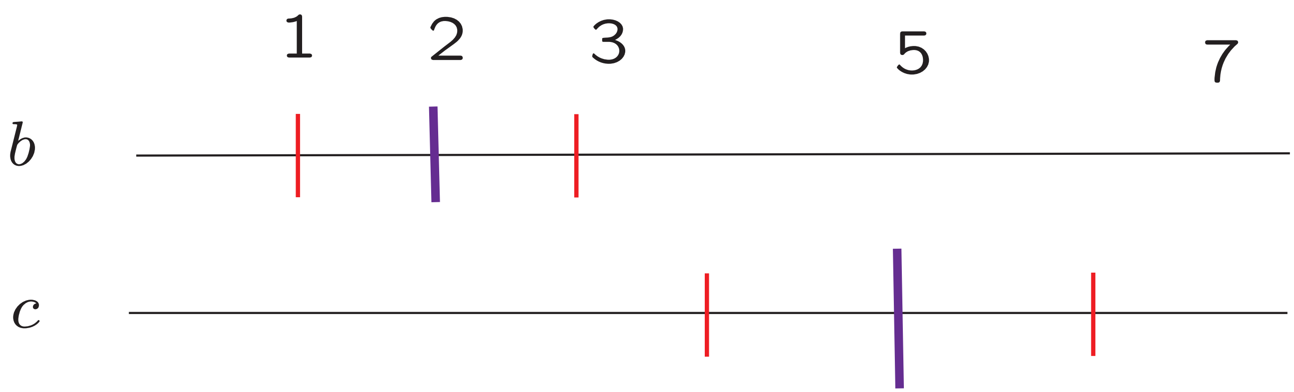
Dann gilt:

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

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

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



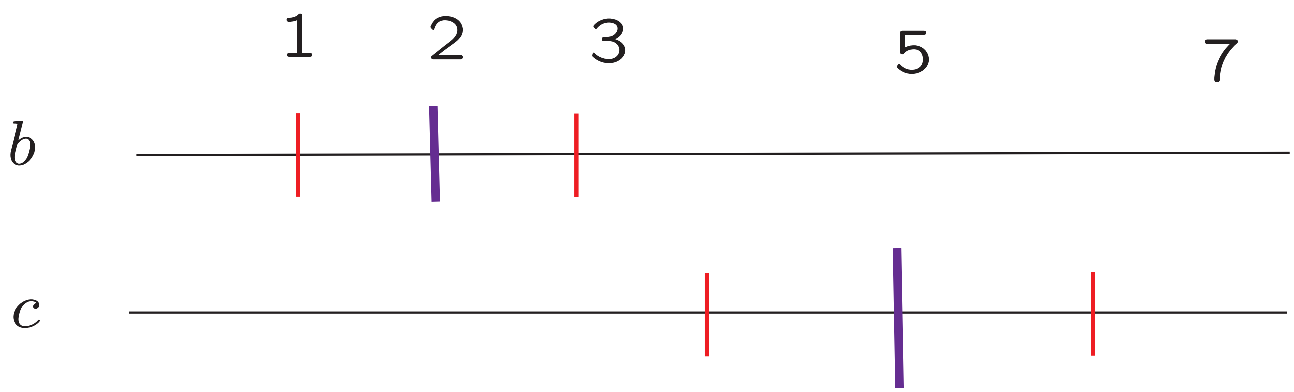
Dann gilt:

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

$\sqrt{\mathbf{Var}_b[X_2]} = 0.9, \quad \sqrt{\mathbf{Var}_c[X_2]} = 1.26.$

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

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



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

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 $\mathbf{E}[N], \mathbf{Var}[N], \mu$ 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:

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

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

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

Zur **Herleitung der Formel von der Zerlegung der Varianz** rekapitulieren wir erst die für jede Zufallsvariable Z (mit $(E[Z^2] < \infty)$ und jede Konstante c geltende Beziehung

$$\mathbf{E}[(Z-c)^2] = \mathbf{Var}[Z] + (\mathbf{E}Z - c)^2 \quad \text{(siehe VL 8b)}$$

Diese überträgt sich auf den *bedingten* Erwartungswert und die *bedingte Varianz*:

Mit $Z := X_2$ folgt

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

Ersetzen wir a_1 durch die Zufallsvariable X_1:

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

und bilden den Erwartungswert, dann bekommen wir

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

also (mit $c := \mathbf{E}[X_2]$)

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

4. Die bedingte Erwartung als beste Prognose im quadratischen Mittel

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:

Wir zerlegen $\mathbf{E}\Big[(X_2 - h(X_1))^2\Big]$ nach X_1:

$$\mathbf{E}\Big[(X_2 - h(X_1))^2\Big] = \mathbf{E}\Big[\mathbf{E}_{X_1}\Big[(X_2 - h(X_1))^2\Big]\Big]$$
$$= \sum_{a_1} \mathbf{P}(X_1 = a_1)\mathbf{E}_{a_1}\Big[(X_2 - h(a_1))^2\Big]$$

Wir wissen schon (aus Vorlesung 8b):

$\mathbf{E}_{a_1}\Big[(X_2 - h(a_1))^2\Big]$ wird minimal für

$h(a_1) := \mathbf{E}_{a_1}[X_2]$. $\square$

Fazit:

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.

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.