

Vorlesung 9a

Bedingte Erwartung und bedingte Varianz

1. Zerlegung eines Erwartungswertes nach der ersten Stufe

(Buch S. 91)

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**:

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

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

Auch der Erwartungswert einer reellwertigen Zufallsvariablen $g(X_1, X_2)$ kann nach der ersten Stufe zerlegt werden.

Sei $h : S_1 \times S_2 \rightarrow \mathbb{R}$.

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

Für $a_1 \in S_1$ setzen wir

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

und nennen diese Zahl den

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

gegeben $\{X_1 = a_1\}$.

Merke:

Der bedingte Erwartungswert

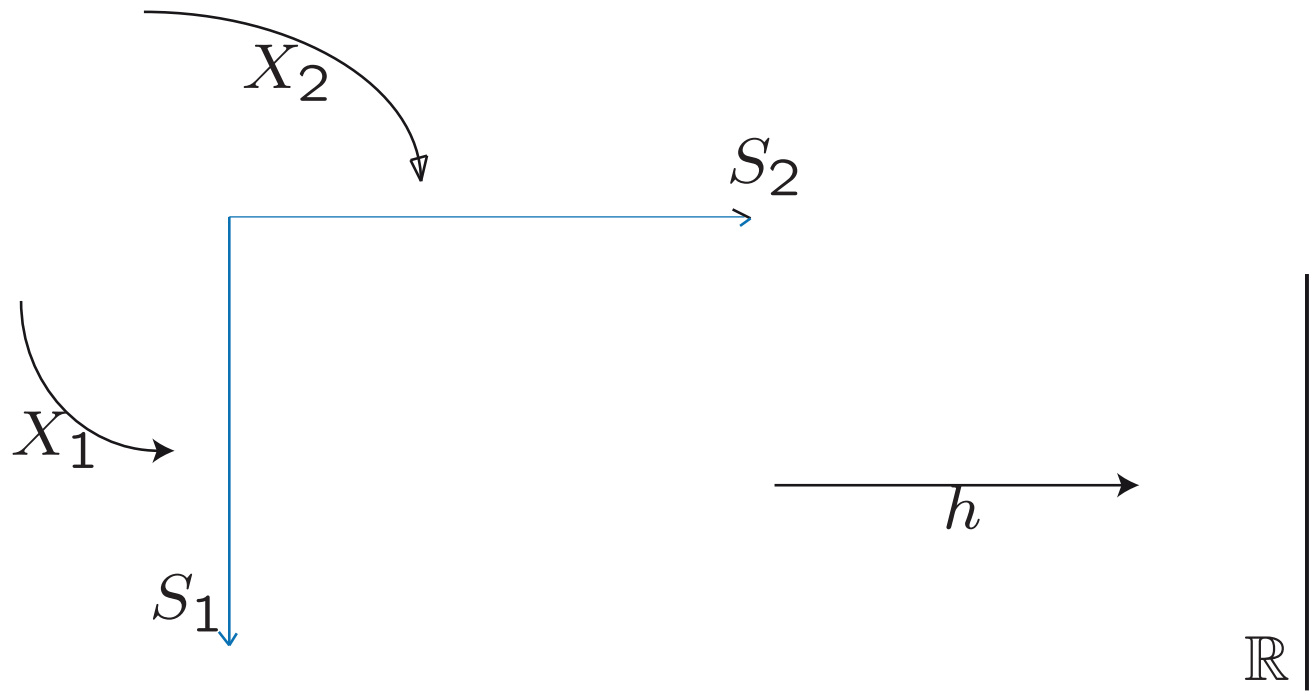
$$\mathbf{E}_{a_1}[h(X_1, X_2)]$$

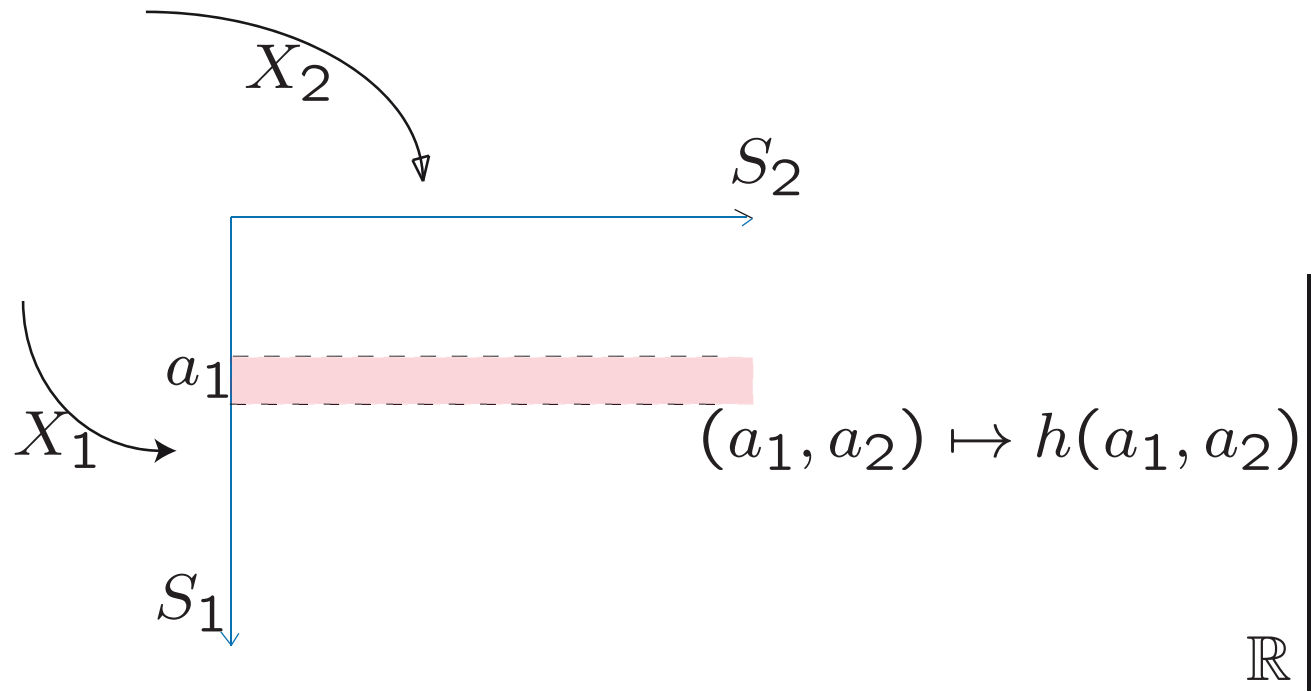
wird gebildet mit der Übergangsverteilung $P(a_1, \cdot)$,

also mit den Wahrscheinlichkeitsgewichten,

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

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





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

(Zerlegung des Erwartungswerts nach der ersten Stufe)

$$\begin{aligned} & \mathbf{E}[h(X_1, X_2)] \\ &= \sum_{a_1 \in S_1} \sum_{a_2 \in S_2} h(a_1, a_2) \mathbf{P}(X_1 = a_1, X_2 = a_2) \\ &= \sum_{a_1 \in S_1} \sum_{a_2 \in S_2} h(a_1, a_2) \mathbf{P}(X_1 = a_1) \mathbf{P}_{a_1}(X_2 = a_2) \end{aligned}$$

$$\begin{aligned}
& \mathbf{E}[h(X_1, X_2)] \\
&= \sum_{a_1 \in S_1} \sum_{a_2 \in S_2} h(a_1, a_2) \mathbf{P}(X_1 = a_1, X_2 = a_2) \\
&= \sum_{a_1 \in S_1} \sum_{a_2 \in S_2} h(a_1, a_2) \mathbf{P}(X_1 = a_1) \mathbf{P}_{a_1}(X_2 = a_2) \\
&= \sum_{a_1 \in S_1} \mathbf{P}(X_1 = a_1) \mathbf{E}_{a_1}[h(a_1, X_2)] \\
&= \mathbf{E}[\mathbf{E}_{X_1}[h(X_1, X_2)]].
\end{aligned}$$

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

$$\begin{aligned} &= \sum_{a_1 \in S_1} \mathbf{P}(X_1 = a_1) \mathbf{E}_{a_1}[h(a_1, X_2)] \\ &= \mathbf{E}[\mathbf{E}_{X_1}[h(X_1, X_2)]]. \end{aligned}$$

“Zerlegung des Erwartungswertes nach der ersten Stufe”

Merke: Der bedingte Erwartungswert

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

ist eine Zahl.

$$\mathbf{E}_{X_1} [h(X_1, X_2)]$$

ist eine **Zufallsvariable**.

Wir nennen diese Zufallsvariable die *bedingte Erwartung* von $h(X_1, X_2)$ gegeben X_1 .

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

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

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

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

Wir haben dann die einprägsame Formel

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

(Zerlegung des Erwartungswertes von X_2 nach X_1 .)

2. Ein Beispiel: Suchen in Listen.

(Buch S. 85-87)

n Namen werden in ℓ Listen einsortiert. Dadurch ergibt sich ein ℓ -Tupel $k = (k_1, \dots, k_\ell)$ von Listenlängen (eine "Besetzung" k der Plätze $1, \dots, \ell$)

Jeder Name steht in seiner Liste Nr. j an einer der Stellen $i = 0, \dots, k_j - 1$.

Dieses i bezeichnet man auch als (Such-)Tiefe des Namens.

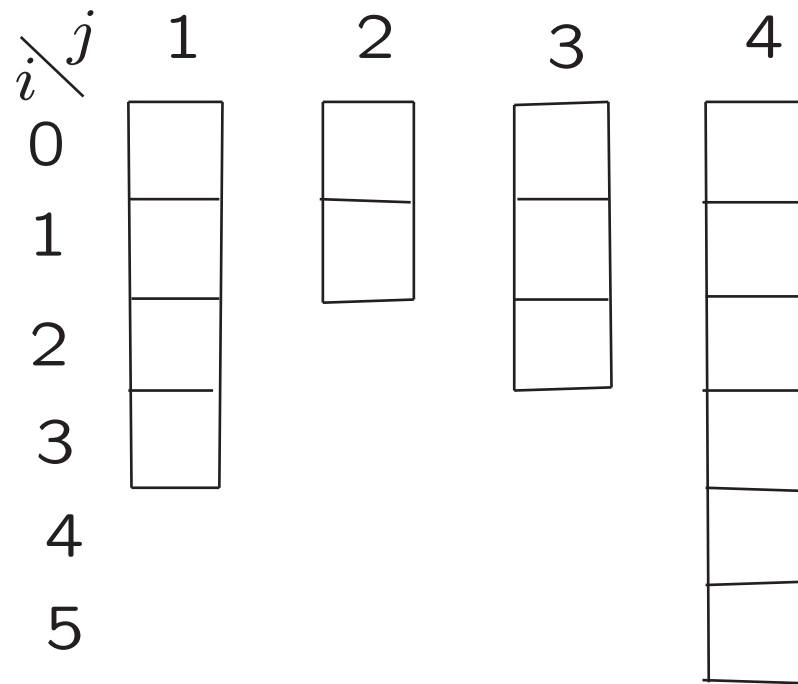
Vorstellung: Die Listennummer $j = 1, \dots, \ell$ entspricht dem Anfangsbuchstaben des Namens.

Beispielsweise ist

für $\ell = 4$ mögliche Anfangsbuchstaben

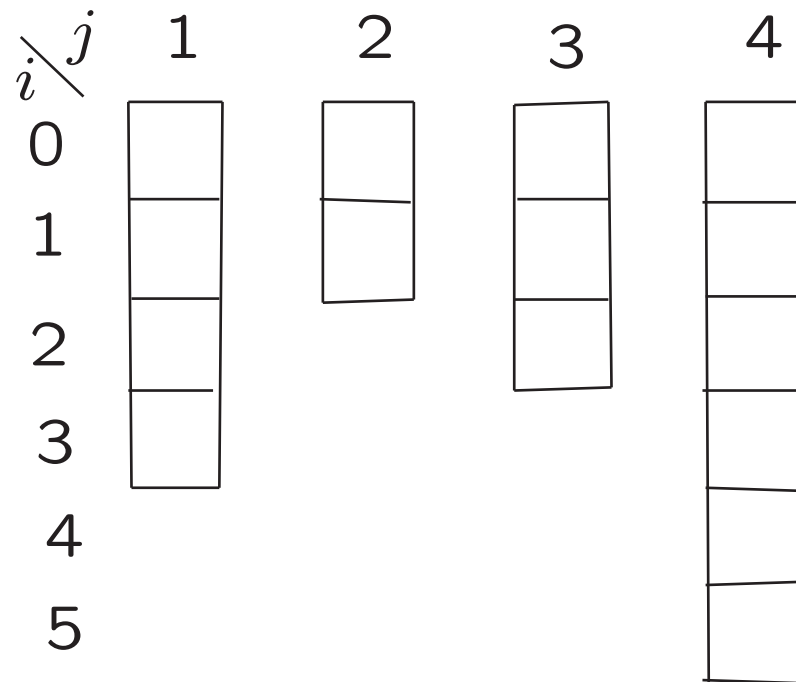
und $n = 15$ Namen

eine mögliche Besetzung:

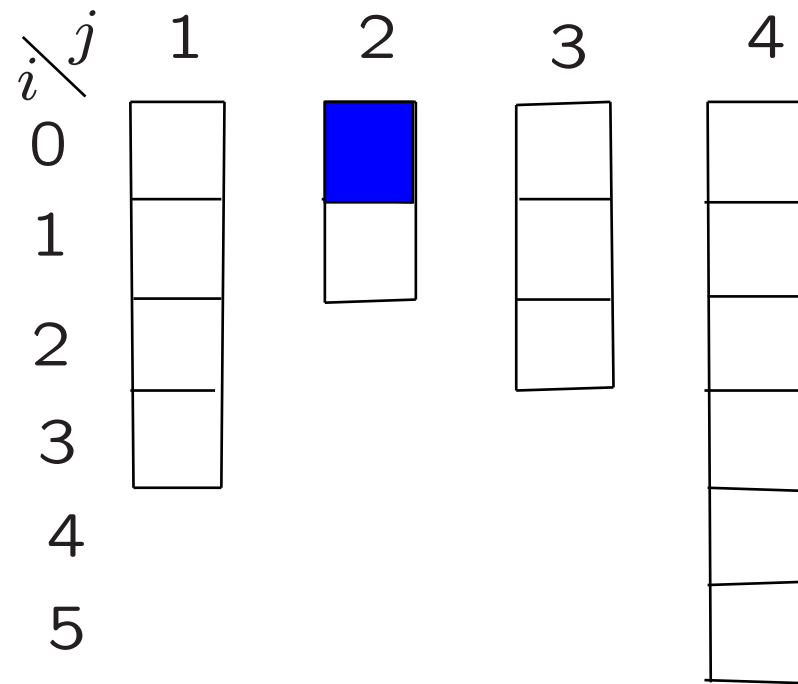


$$n = 15, \ell = 4$$

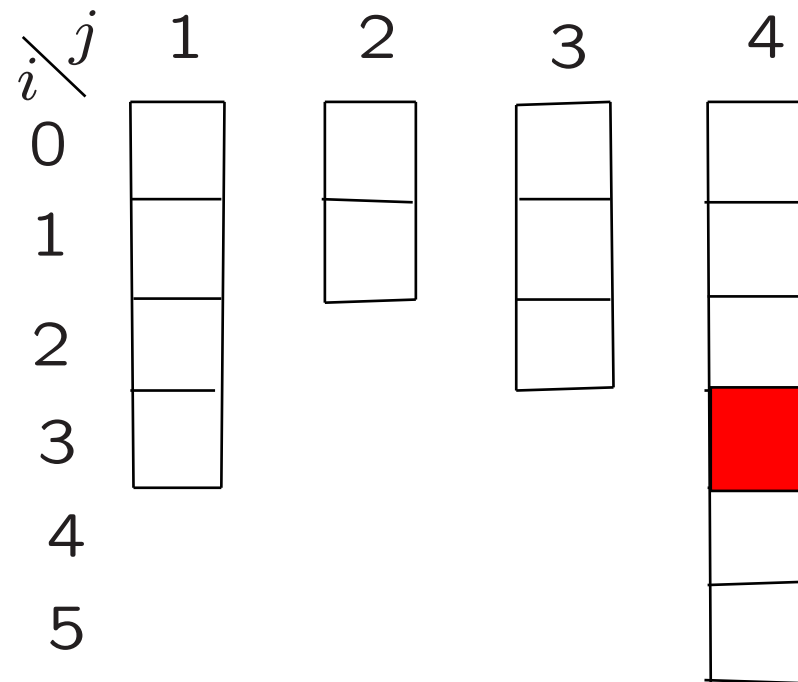
$$k = (k_1, k_2, k_3, k_4) = (4, 2, 3, 6)$$



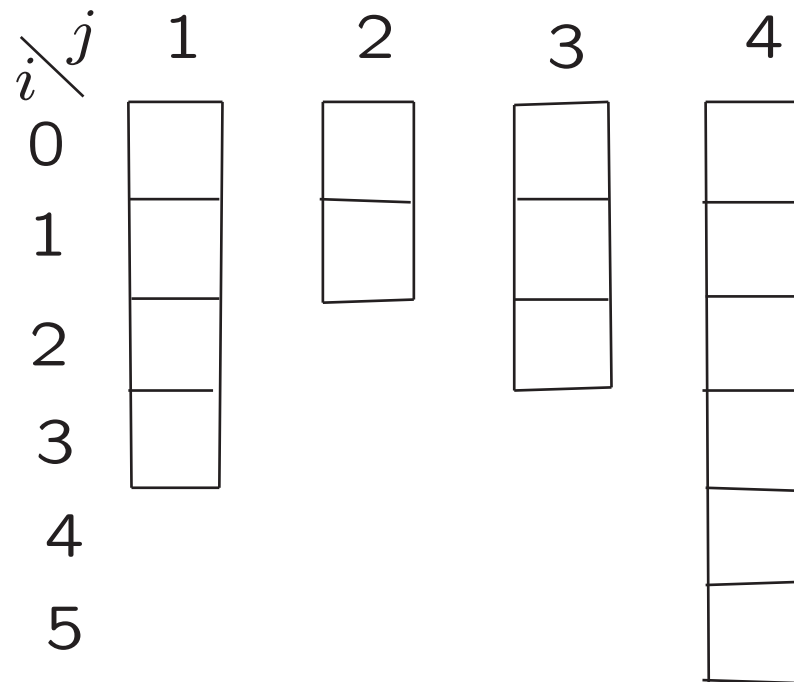
Erste Frage: Was ist für gegebene Listenlängen $k = (k_j)$
 der Erwartungswert der Tiefe T
 eines rein zufällig aus den n herausgegriffenen Names?



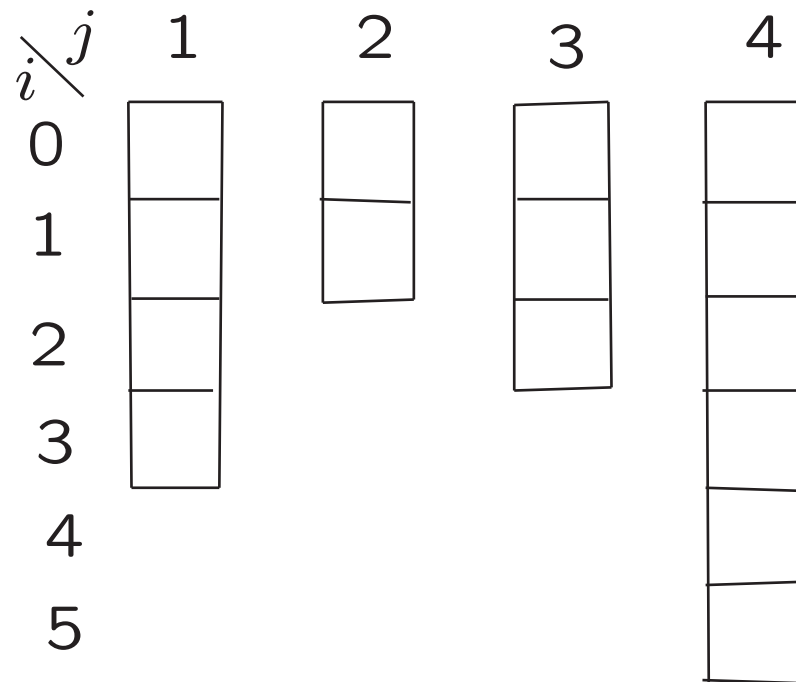
Liste $j = 2$,
Tiefe $i = 0$.



Liste $j = 4$,
Tiefe $i = 3$.



Was ist *bei gegebenem k*
 der Erwartungswert der Suchtiefe T
 eines rein zufällig aus den n herausgegriffenen Names?



Die Antwort ist

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

Wir betrachten jetzt ein
stochastisches Modell für die erste Stufe:

Annahme:

Die zufällige Besetzung $Z = (Z_1, \dots, Z_\ell)$

kommt durch n -maliges Würfeln

mit den Gewichten p_1, \dots, p_ℓ zustande.

Z ist multinomial (n, p_1, \dots, p_ℓ) -verteilt.

(Vorstellung: Die n Namen

sind eine Stichprobe aus einer großen Population
mit bekannten Anteilen der Anfangsbuchstaben.)

Aus den n Namen wird rein zufällig einer herausgegriffen.

Es sei T die Tiefe des Namens in seiner Liste.

Aufgabe: Berechne $E[T]$.

Diese Erwartungswert beschreibt die mittlere Suchzeit

(Suchtiefe)

aller in den Listen vorhandenen Namen.

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

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

Mit der oben hergeleiteter Zerlegung des Erwartungswertes

$$\mathbf{E}[T] = \mathbf{E}[\mathbf{E}_Z[T]]$$

erhalten wir

$$\mathbf{E}[T] = \frac{1}{n} \sum_{j=1}^{\ell} \mathbf{E}\left[\frac{Z_j(Z_j - 1)}{2}\right] .$$

$$\mathbf{E}[T] = \frac{1}{n} \sum_{j=1}^{\ell} \mathbf{E}\left[\frac{Z_j(Z_j - 1)}{2}\right]$$

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

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

$$\begin{aligned} \mathbf{E}[Z_j(Z_j - 1)] &= \text{Var}[Z_j] + \mathbf{E}[Z_j]^2 - \mathbf{E}[Z_j] \\ &= np_j(1 - p_j) + (np_j)^2 - np_j = p_j^2 n(n - 1), \end{aligned}$$

$$\mathbf{E}[T] = \frac{n-1}{2} (p_1^2 + \dots + p_\ell^2).$$

$$\mathbf{E}[T] = \frac{n-1}{2}(p_1^2 + \dots + p_\ell^2).$$

Im Fall uniformer Gewichte

$$p_1 = \dots = p_\ell = 1/\ell$$

ergibt sich

$$\mathbf{E}[T] = \frac{n-1}{2\ell}.$$

Eine Modifikation des vorigen Beispiels:

Sei Z wieder multinomial (n, p_1, \dots, p_ℓ) -verteilt,
 J sei unabhängig von Z , mit $\mathbf{P}(J = j) = p_j, \quad j = 1, \dots, \ell.$

Berechne den Erwartungswert von Z_J .

(Dieser Erwartungswert beschreibt die mittlere Suchzeit nach einem in den Listen nicht vorhandenen Namen.)

Wir fassen Z als erste Stufe auf, und Z_J als zweite, und ...

... zerlegen $\mathbf{E}[Z_J]$ nach den Ausgängen von Z :

$$\begin{aligned}\mathbf{E}[Z_J] &= \sum_k \mathbf{P}(Z = k) \mathbf{E}_k[Z_J] \\ &= \sum_k \mathbf{P}(Z = k) \mathbf{E}_k[k_J] = \sum_k \mathbf{P}(Z = k) \sum_{j=1}^{\ell} k_j p_j \\ &= \sum_{j=1}^{\ell} p_j \sum_k k_j \mathbf{P}(Z = k) \\ &= \sum_{j=1}^{\ell} p_j \mathbf{E}Z_j = \sum_{j=1}^{\ell} p_j np_j = n \sum_{j=1}^{\ell} p_j^2.\end{aligned}$$

Im Fall uniformer Gewichte

$$p_1 = \dots = p_\ell = 1/\ell$$

ergibt sich

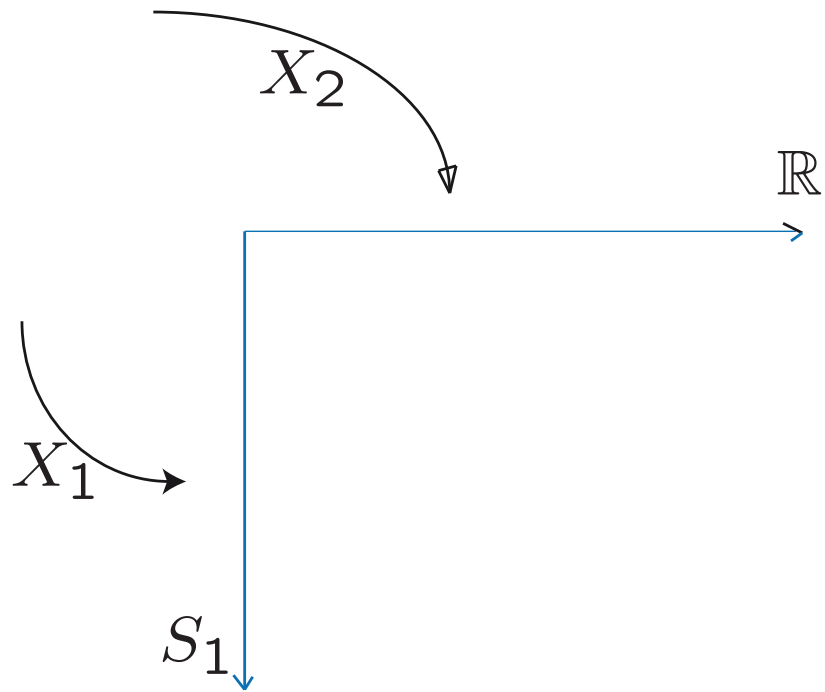
$$\mathbf{E}[Z_J] = \frac{n}{\ell}.$$

Im Vergleich dazu war (siehe voriges Beispiel)
die mittlere Suchtiefe eines rein zufällig aus den n
herausgegriffenen Namens

$$\mathbf{E}[T] = \frac{n-1}{2\ell}.$$

3. Bedingte Varianz

(Buch S. 90)



Für reellwertiges X_2 definieren wir die
bedingte Varianz von X_2 , gegeben $\{X_1 = a_1\}$:

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

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

Hilfreiche Formel für die Varianz von $Y - c$:

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

Daraus folgt (für alle $a_1 \in S_1$ und beliebiges $h(a_1) \in \mathbb{R}$):

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

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

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 die Formel von dessen Zerlegung nach X_1 :

$$\begin{aligned} (*) \quad & \mathbf{E}[(X_2 - h(X_1))^2] \\ &= \mathbf{E}[\mathbf{Var}_{X_1}[X_2]] + \mathbf{E}[(\mathbf{E}_{X_1}[X_2] - h(X_1))^2]. \end{aligned}$$

4. 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

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

Beweis:

$$\begin{aligned} (*) \quad & \mathbf{E}[(X_2 - h(X_1))^2] \\ &= \mathbf{E}[\text{Var}_{X_1}[X_2]] + \mathbf{E}[(\mathbf{E}_{X_1}[X_2] - h(X_1))^2]. \end{aligned}$$

Der Satz über die Positivität des Erwartungswerts impliziert:

Der letzte Term in (*) rechts wird minimal
(nämlich 0) genau dann, wenn

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

Äquivalent dazu:

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

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.

5. Zerlegung der Varianz

(Buch S. 90)

$$\begin{aligned} (*) \quad & \mathbf{E}[(X_2 - h(X_1))^2] \\ &= \mathbf{E}[\mathbf{Var}_{X_1}[X_2]] + \mathbf{E}[(\mathbf{E}_{X_1}[X_2] - h(X_1))^2]. \end{aligned}$$

$$\begin{aligned} (*) \quad & \mathbf{E}[(X_2 - h(X_1))^2] \\ &= \mathbf{E}[\mathbf{Var}_{X_1}[X_2]] + \mathbf{E}[(\mathbf{E}_{X_1}[X_2] - h(X_1))^2]. \end{aligned}$$

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

$$\begin{aligned} (*) \quad & \mathbf{E}[(X_2 - h(X_1))^2] \\ &= \mathbf{E}[\mathbf{Var}_{X_1}[X_2]] + \mathbf{E}[(\mathbf{E}_{X_1}[X_2] - h(X_1))^2]. \end{aligned}$$

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

dann wird der **zweite Term rechts in (*)**

$$\begin{aligned}
 (*) \quad & \mathbf{E}[(X_2 - h(X_1))^2] \\
 &= \mathbf{E}[\text{Var}_{X_1}[X_2]] + \mathbf{E}[(\mathbf{E}_{X_1}[X_2] - \mathbf{E}[X_2])^2].
 \end{aligned}$$

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

dann wird der **zweite Term rechts in (*)** wegen

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

zu

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

Insgesamt wird dann die Formel (*) zu

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

Zerlegung der Varianz von X_2 nach X_1 .

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

Zum Merken:

Die **Varianz von X_2**

ist die Summe aus dem

Erwartungswert der **bedingten Varianzen**

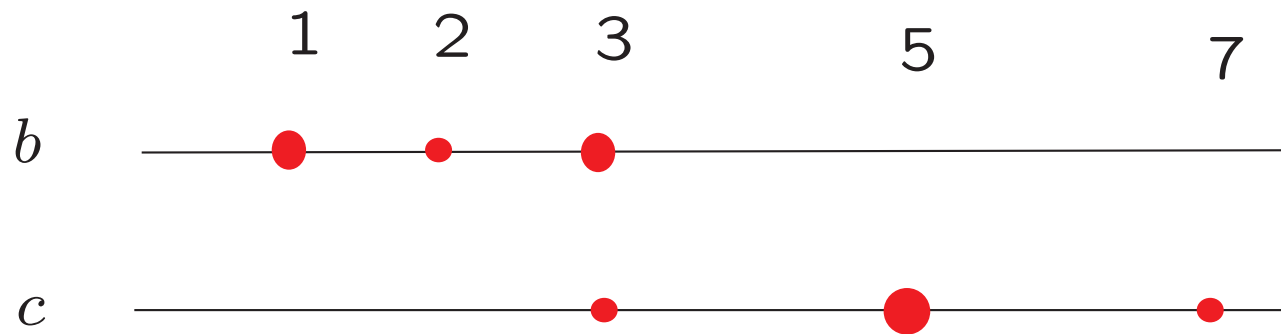
und der **Varianz** der **bedingten Erwartungswerte**

(Variabilität **innerhalb** der Zeilen
plus Variabilität **zwischen** den Zeilen).

Wir illustrieren dies mit einem kleinen Beispiel:

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



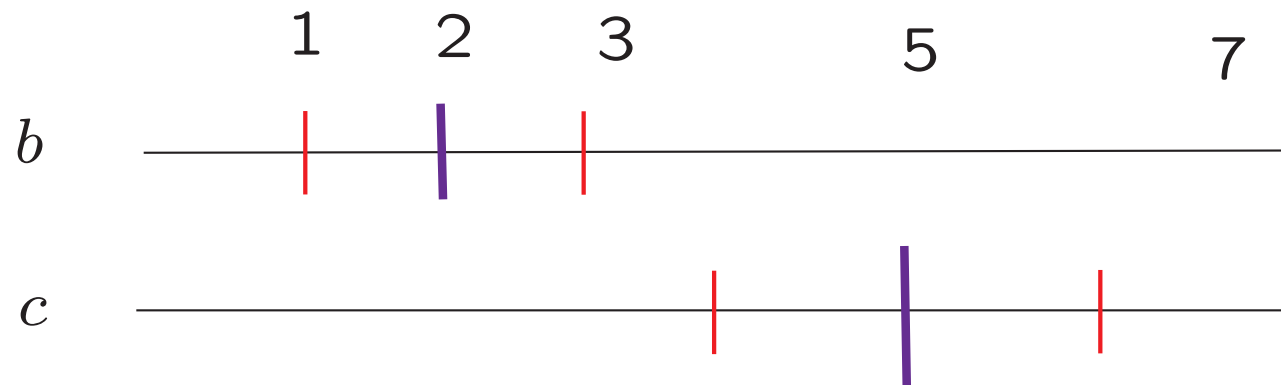
Dann gilt:

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

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

$$E_b[X_2] = 2, E_c[X_2] = 5,$$

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



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

Wir erklären noch das Ergebnis

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

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, \dots 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 .

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

Wir nehmen N als erste und Y als zweite Stufe:

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

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

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