Algorithmen und Wahrscheinlichkeit

Woche 7

Ilya Maier

Varianz

Sei X ein Zufallsvariable mit $\mu = \mathbb{E}[X]$

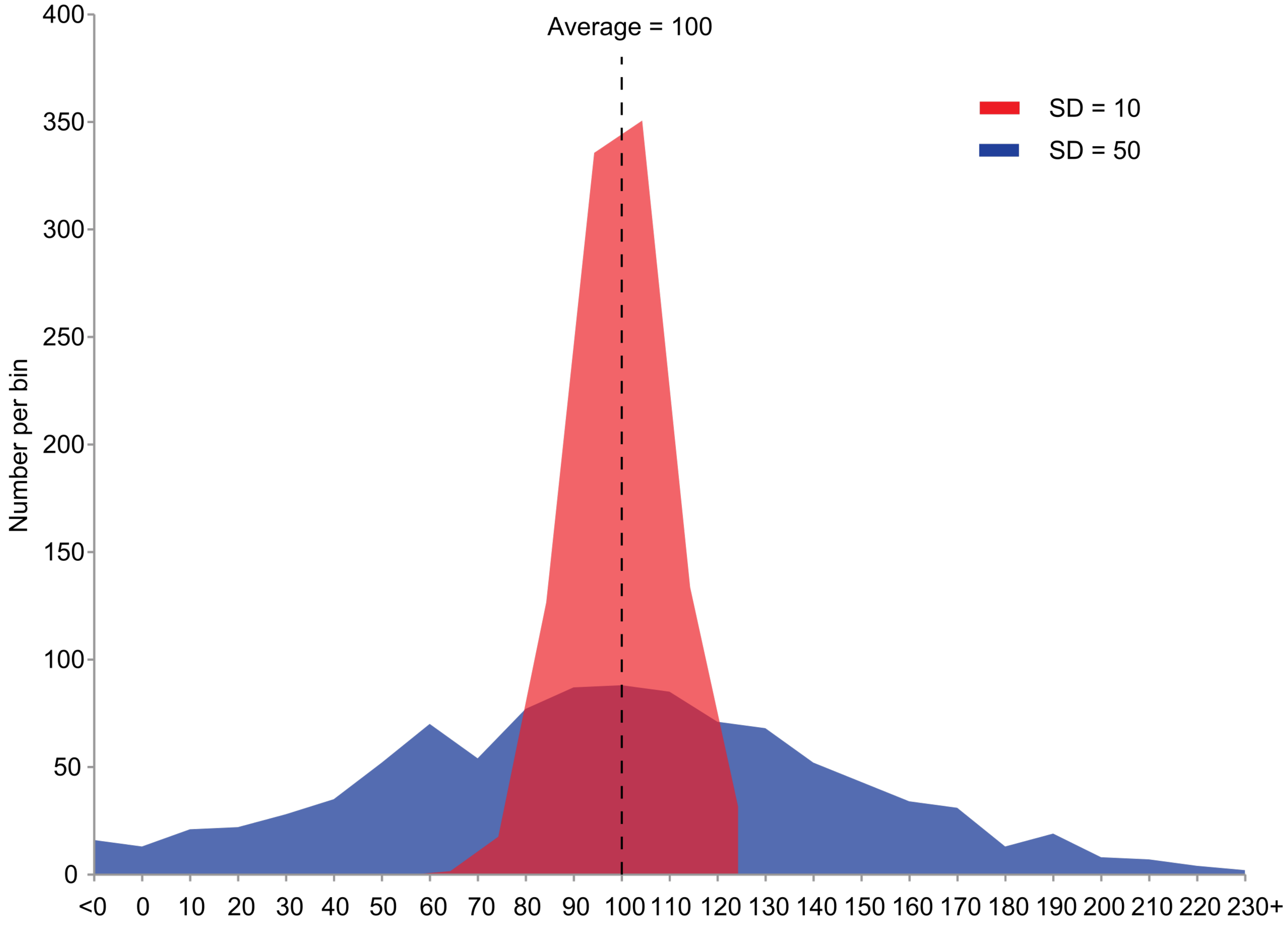
Varianz: $\text{Var}[X] := \mathbb{E}[(X-\mu)^2] = \sum_{x \in W_X} (x-\mu)^2 \cdot \Pr[X = x]$

Standardabweichung von X: $\sigma = \sqrt{\text{Var}[X]}$

Varianz mit Erwartungswert: $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

Rechenregeln:

1) $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ $\quad \forall X, Y$
2) $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ $\quad \forall X, Y$ unabhängig
3) $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ $\quad \forall X, Y$ unabhängig
4) $\text{Var}[X \cdot Y] \neq \text{Var}[X] \cdot \text{Var}[Y]$ $\quad$ (in meisten Fällen)
5) $\text{Var}[a \cdot X + b] = a^2 \cdot \text{Var}[X]$ $\quad \forall a, b \in \mathbb{R}$



Abschätzen von Wahrscheinlichkeiten

Markovs Ungleichung:

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

$\forall X \geq 0$

$\forall t > 0,\ t \in \mathbb{R}$

Chebyshevs Ungleichung:

$$\Pr[\,|X - \mathbb{E}[X]|\, \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}$$

$\forall X$

$\forall t > 0,\ t \in \mathbb{R}$

Chernoffs Ungleichung:

1. $\Pr[X \geq (1+\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{3}\delta^2\mathbb{E}[X]}$
2. $\Pr[X \leq (1-\delta)\mathbb{E}[X]] \leq e^{-\frac{1}{2}\delta^2\mathbb{E}[X]}$
3. $\Pr[X \geq t] \leq 2^{-t}$

$\forall X \sim \mathrm{Bin}(n, p)$

$\forall 0 < \delta \leq 1$

$\forall t \geq 2e\mathbb{E}[X]$

Randomisierte Algorithmen

Randomisierter Algorithmus : Eingabe $I \rightarrow$ Algorithmus A mit Zufallszahlen $R \rightarrow$ Ausgabe $A(I, R)$

- deterministisch: selbe Eingabe, selber Output
- nicht-deterministisch: selbe Eingabe, nicht unbedingt selber Output

Monte Carlo Algorithmus: Primzahltest, Target-shooting
→ <u>Korrektheit</u> ist die Zufallsvariable

- immer gleiche Laufzeit
- manchmal falsches Ergebnis

Las Vegas Algorithmus: Quicksort, Duplikate finden
→ <u>Laufzeit</u> ist die Zufallsvariable
→ Geometrisch verteilt mit $p = \Pr[A(I) \neq \text{"???"}]$

- immer korrekte Antwort
- manchmal dauert zu lange /
  gibt nach einer bestimmter Zeit “???” aus

Fehlerreduktion

Las-Vegas:

Sei A ein Las-Vegas-Algorithmus mit $\Pr[A(I)$ ist korrekt$] \geq \epsilon$

Sei A_δ für $\delta > 0$ ein Algorithmus, der entweder die erste Ausgabe verschieden von $???$ ausgibt

oder der nach $N = \lceil \epsilon^{-1} \cdot \ln(\delta^{-1}) \rceil$ Versuchen $???$ ausgibt

dann gilt $\Pr[A_\delta(I)$ ist falsch$] \leq \delta$

ϵ	δ	N
0.1	0.01	47
0.5	0.01	10
0.5	10^{-80}	369
0.9	10^{-30}	77

Fehlerreduktion

Monte-Carlo - Einseitiger Fehler:

$\Pr[A(I) = \text{Ja}] = 1$ für alle Ja-Instanzen I
$\Pr[A(I) = \text{Nein}] \geq \epsilon$ für alle Nein-Instanzen I

$\implies$ Wenn $A(I) = \text{Ja}$, dann könnte die Ausgabe falsch sein
Wenn $A(I) = \text{Nein}$, dann ist die Ausgabe immer korrekt

Sei A_δ für $\delta > 0$ ein Algorithmus, der entweder Nein ausgibt, sobald das erste Mal Nein vorkommt,

oder der nach $N = \lceil \epsilon^{-1} \cdot \ln(\delta^{-1}) \rceil$ Versuchen Ja ausgibt

dann gilt:

$\Pr[A_\delta(I) = \text{Ja}] = 1$ für alle Ja-Instanzen I

$\Pr[A_\delta(I) = \text{Nein}] \geq 1 - \delta$ für alle Nein-Instanzen I

Monte-Carlo - Zweiseitiger Fehler:

$\Pr[A(I) \text{ ist korrekt}] \geq 0.5 + \epsilon$ für alle Instanzen I

A_δ gibt die meiste Antwort aus nach N Wiederholungen

$\implies$

Für $\delta > 0$, wenn $N = \lceil 4 \cdot \epsilon^{-2} \cdot \ln(\delta^{-1}) \rceil$

$\Pr[A_\delta(I) \text{ ist falsch}] \leq \delta$ für alle Instanzen I

Kahoot

Aufgaben