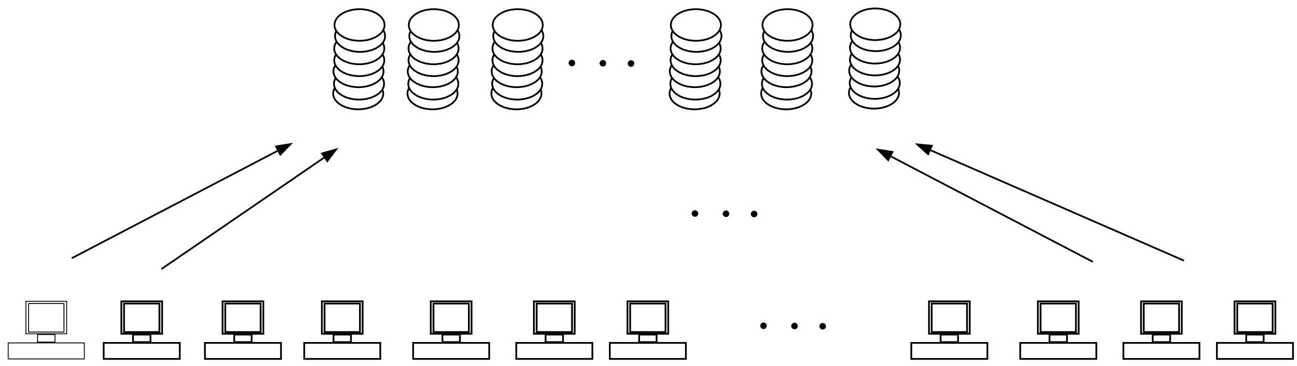
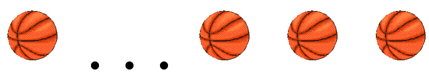
BSP.: Fallbeispiel 1: Zuordnung von Datenbankanfragen

Szenario:



Aufgabe: Entwerfe ein möglichst einfaches Verfahren, das jede Anfrage genau einem Server zuordnet — und zwar so, daß alle Server möglichst gleich viel Arbeit erhalten.

Idealisiertes Modell: Wir werfen n Bälle *zufällig* in m Urnen.



$X_i :=$ Anzahl Bälle in $i$ter Urne

Frage: Wie groß ist $\max\limits_{1\le i\le n} X_i$?

[Wir betrachten im folgenden nur den Fall $n = m$.]

$X_i :=$ Anzahl Bälle in $i$ter Urne

Frage: Wie groß ist $\max\limits_{1 \le i \le n} X_i$?

Antwort: Mit hoher Wahrscheinlichkeit etwa $\dfrac{\ln n}{\ln \ln n}$.

Beweisidee: Wir zeigen

$$\Pr[\max_{1 \le i \le n} X_i \ge \alpha \frac{\ln n}{\ln \ln n}] = \begin{cases} o(1) & \text{, falls } \alpha > 1 \\ 1 - o(1) & \text{, falls } \alpha < 1. \end{cases}$$

Setze

$$Y_i := \begin{cases} 1 & \text{, falls } X_i \ge \alpha \frac{\ln n}{\ln \ln n} \\ 0 & \text{, sonst} \end{cases}$$

Dann gilt für $Y := \sum_{i=1}^{n} Y_i$:

$$Y = 0 \qquad \Longleftrightarrow \qquad \max_{1 \le i \le n} X_i < \alpha \frac{\ln n}{\ln \ln n}$$

Markov-Ungleichung: $\Pr[Y \ge 1] \le \mathbb{E}[Y]$

Chebychev-Ungleichung: $\Pr[Y = 0] \le \dfrac{\mathrm{Var}[Y]}{(\mathbb{E}[Y])^2}$

Etwas Rechnung (weggelassen):

$$\begin{aligned}\mathbb{E}[Y] &\approx n^{1-\alpha}\\ \mathrm{Var}[Y] &\approx \mathbb{E}[Y]\end{aligned}$$

Also

$$\begin{aligned}\Pr[\max_{1\le i\le n} X_i \ge \alpha\frac{\ln n}{\ln\ln n}] &= \Pr[Y \ge 1] \le \mathbb{E}[Y]\\ &= o(1), \qquad \text{falls } \alpha > 1\end{aligned}$$

$$\begin{aligned}\Pr[\max_{1\le i\le n} X_i < \alpha\frac{\ln n}{\ln\ln n}] &= \Pr[Y = 0]\\ &\le \frac{\mathrm{Var}[Y]}{(\mathbb{E}[Y])^2}\\ &\approx \frac{1}{\mathbb{E}[Y]}\\ &= o(1), \qquad \text{falls } \alpha < 1\end{aligned}$$

Satz: (*Chernoff-Schranken*)

Seien $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit $\Pr[X_i = 1] = p_i$ und $\Pr[X_i = 0] = 1 - p_i$. Dann gilt für $X := \sum_{i=1}^n X_i$ und $\mu := \mathbb{E}[X] = \sum_{i=1}^n p_i$, sowie jedes $\delta > 0$, dass

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}} \right)^\mu .$$

Weiter gilt für alle $0 < \delta < 1$, dass

$$\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}} \right)^\mu .$$

Satz: (*Chernoff-Schranken — „Arbeitsversionen“*)

Seien $X_1, \ldots, X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit $\Pr[X_i = 1] = p_i$ und $\Pr[X_i = 0] = 1 - p_i$. Dann gelten folgende Ungleichungen für $X := \sum_{i=1}^n X_i$ und $\mu := \mathbb{E}[X] = \sum_{i=1}^n p_i$:

1. $\Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3}$ für alle $0 < \delta \leq 1$,
2. $\Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}$ für alle $0 < \delta \leq 1$,
3. $\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$ für alle $0 < \delta \leq 1$,
4. $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu}$ und
5. $\Pr[X \geq t] \leq 2^{-t}$ für $t \geq 2e\mu$.

Beweis: Folgt mit Hilfe geeigneter Abschätzungen aus der Analysis unmittelbar aus den ursprünglichen Chernoff-Schranken.

Definition: (*kontinuierliche Zufallsvariable*)

Eine *kontinuierliche* oder auch *stetige Zufallsvariable* X und ihr zugrunde liegender *kontinuierlicher (reeller) Wahrscheinlichkeitraum* sind definiert durch eine integrierbare Dichte(-funktion) $f_X : \mathbb{R} \to \mathbb{R}_0^+$ mit der Eigenschaft

$$\int_{-\infty}^{+\infty} f_X(x) \, \mathrm{d}x = 1.$$

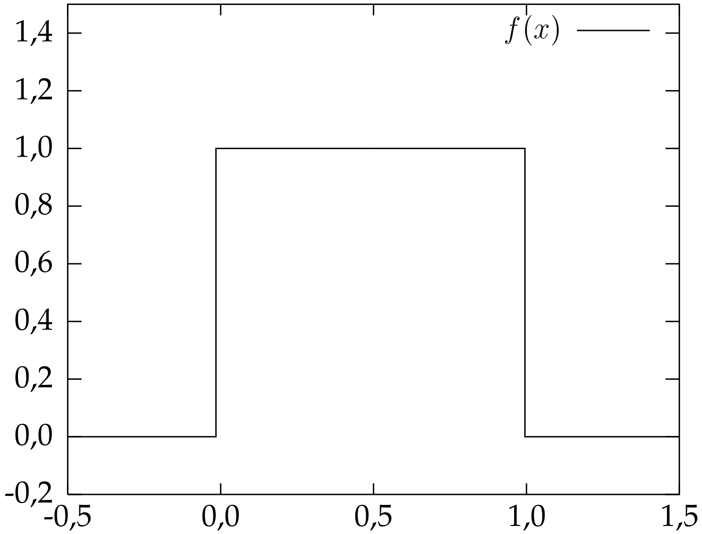
Eine Menge $A \subseteq \mathbb{R}$, die durch Vereinigung $A = \bigcup_k I_k$ abzählbar vieler paarweise disjunkter Intervalle beliebiger Art (offen, geschlossen, halboffen, einseitig unendlich) gebildet werden kann, heißt *Ereignis*. Ein Ereignis A tritt ein, wenn X einen Wert aus A annimmt. Die Wahrscheinlichkeit von A ist bestimmt durch

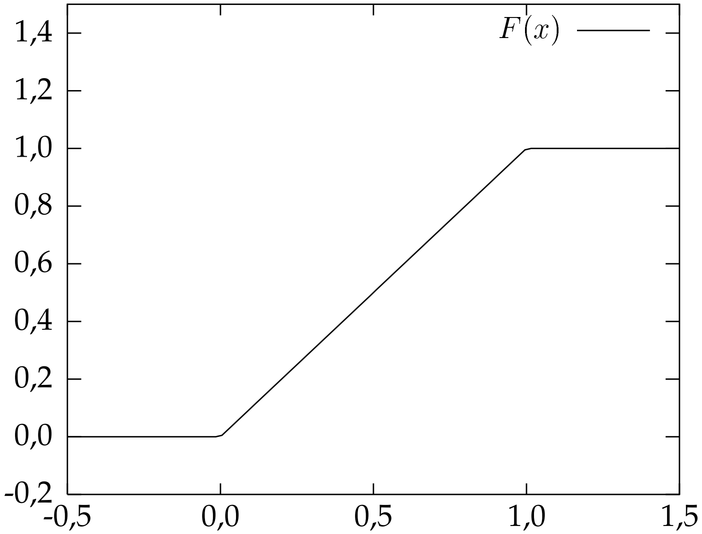
$$\Pr[A] = \int_A f_X(x) \, \mathrm{d}x = \sum_k \int_{I_k} f_X(x) \, \mathrm{d}x.$$

Gleichverteilung auf $[a,b]$:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } x \in [a,b], \\ 0 & \text{sonst.} \end{cases}$$

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \begin{cases} 0 & \text{für } x < a, \\ \frac{x-a}{b-a} & \text{für } a \le x \le b, \\ 1 & \text{für } x > b. \end{cases}$$



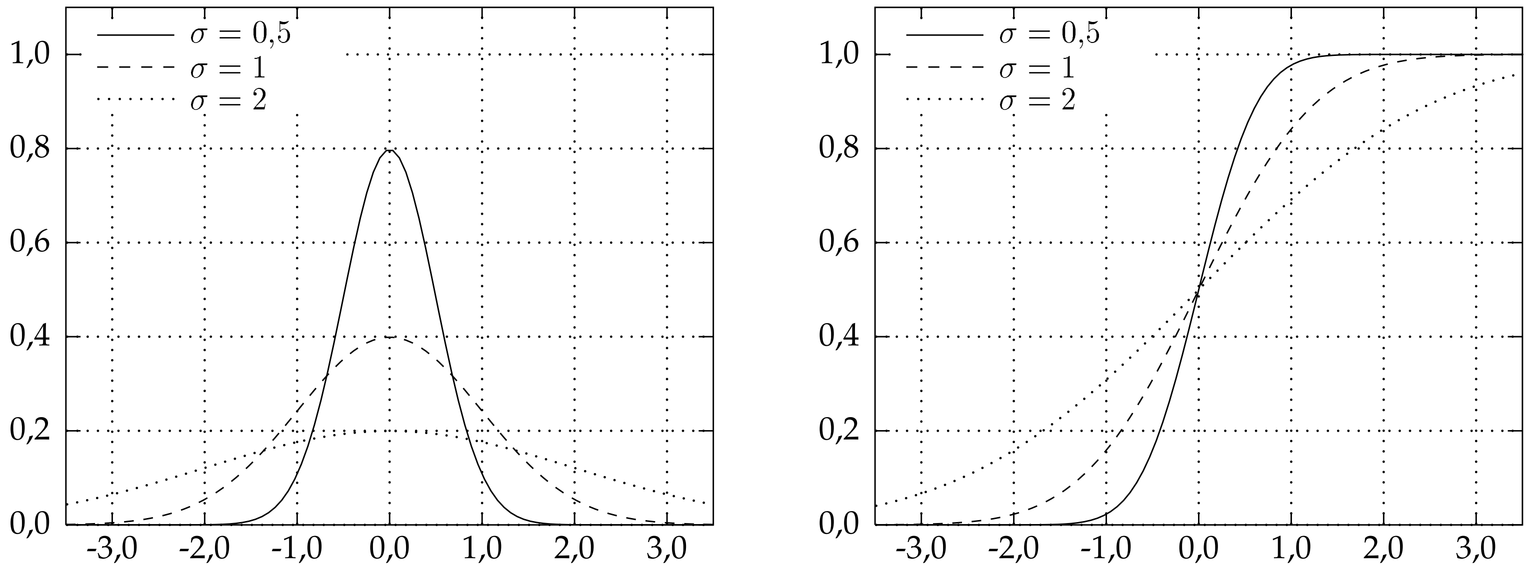


Gleichverteilung über dem Intervall $[0, 1]$

Normalverteilung mit Parameter (μ, σ):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) =: \phi(x;\mu,\sigma)$$

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)\, \mathrm{d}t =: \Phi(x;\mu,\sigma).$$

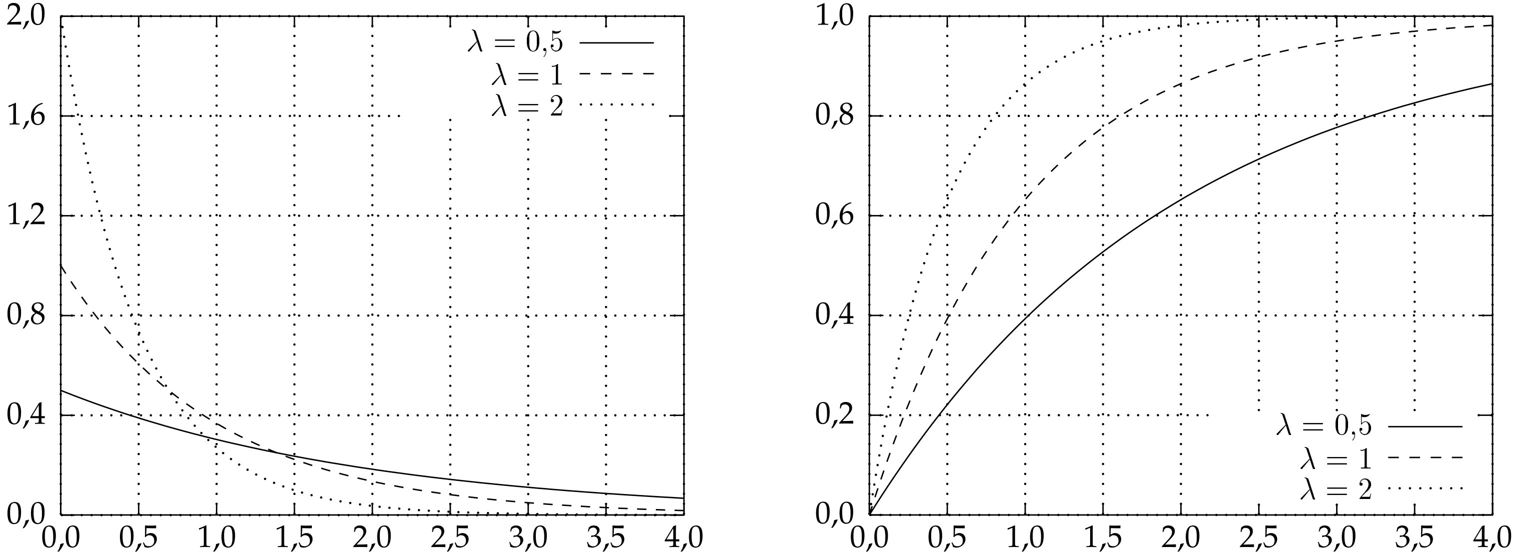


Dichte und Verteilung von $\mathcal{N}(0, \sigma^2)$

Exponentialverteilung mit Parameter λ:

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{falls } x \geq 0, \\ 0 & \text{sonst} \end{cases}$$

$$F(x) = \int_0^x \lambda \cdot e^{-\lambda t} \, \mathrm{d}t = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x} \qquad \text{für } x > 0.$$



Dichte und Verteilung der Exponentialverteilung

Erwartungswert und Varianz

Für eine kontinuierliche Zufallsvariable X mit Dichte $f_X(t)$ ist der Erwartungswert definiert durch

$$\mathbb{E}[X] \quad = \quad \int_{-\infty}^{\infty} t \cdot f_X(t)\ \mathrm{d}t,$$

sofern das Integral $\int_{-\infty}^{\infty} |t| \cdot f_X(t)\ \mathrm{d}t$ endlich ist.

Für die Varianz gilt entsprechend

$$\mathrm{Var}[X] \quad = \quad \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (t - \mathbb{E}[X])^2 \cdot f_X(t)\ \mathrm{d}t,$$

wenn $\mathbb{E}[(X - \mathbb{E}[X])^2]$ existiert.

Beispiele:

- Gleichverteilung auf $[a, b]$:

$$\begin{aligned}
\mathbb{E}[X] &= \int_a^b t \cdot \frac{1}{b-a}\,\mathrm{d}t = \ldots = \frac{a+b}{2}, \\
\mathbb{E}[X^2] &= \frac{1}{b-a} \cdot \int_a^b t^2 \cdot \,\mathrm{d}t = \ldots = \frac{b^2+ba+a^2}{3}, \\
\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \ldots = \frac{(a-b)^2}{12}.
\end{aligned}$$

- Exponentialverteilung mit Parameter λ:

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty t \cdot \lambda \cdot e^{-\lambda t}\,\mathrm{d}t = \ldots = \frac{1}{\lambda}. \\
\mathbb{E}[X^2] &= \int_0^\infty t^2 \cdot \lambda \cdot e^{-\lambda t}\,\mathrm{d}t = \ldots = \frac{2}{\lambda^2} \\
\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{\lambda^2}.
\end{aligned}$$

Laplace-Prinzip in kontinuierlichen Wahrscheinlichkeitsräumen

BEISPIEL: *Bertrand'sches Paradoxon*

Betrachten einen Kreis mit einem eingeschriebenen gleichseitigen Dreieck. Bestimme die Wahrscheinlichkeit, mit der die Länge einer zufällig gewählte Sehne die Seitenlänge dieses Dreiecks übersteigt (Ereignis A).