3 Verteilungsfunktionen (Teil 1)

3.1 Erzeugung von Zufallsgrößen (LK 8.2)

3.1.1 Problemstellung

Messung

Gegeben: Stichproben (=Realisierungen) X_i der ZV X

Gesucht: VF F(x)

Simulation

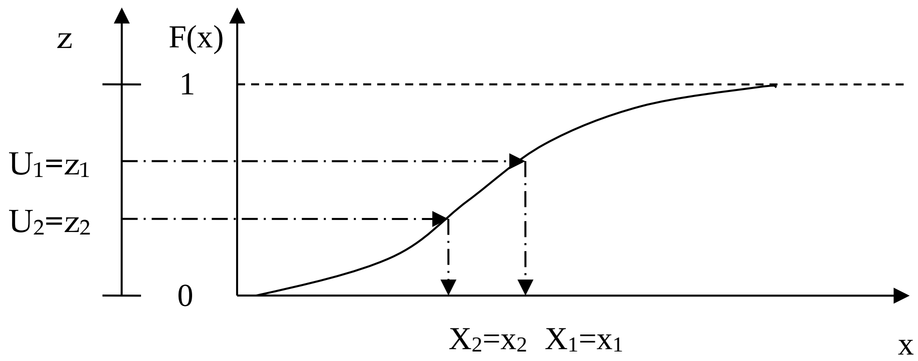
Gegeben: VF F(x)

Gesucht: Erzeugung von Stichproben X_i der ZV X

3.1.2 Prinzipielle Vorgehensweise

- Erzeugung von U(0,1) verteilten Zufallszahlen durch einen Zufallszahlengenerator (z.B. Pseudozufallszahlen durch Kongruenzmethoden, siehe Kapitel 17)
- Transformation dieser Zufallszahlen in gewünschte Stichprobe mit VF F(x)

3.1.3 Inversionsmethode (LK 8.2.1)



- Realisierung einer Zufallszahl $z_i \sim U(0,1)$
- Transformation von z_i in eine Stichprobe X_i gemäß der VF F(x):

$$\boxed{z_i = F(x_i) \quad \rightarrow \quad x_i = F^{-1}(z_i)}$$

- Beispiel: Exponentielle VF
  - $z_i = F(x_i) = 1 - e^{-\lambda \cdot x_i} \quad \rightarrow \quad x_i = -\frac{1}{\lambda}\ln(1-z_i) = -\frac{1}{\lambda}\ln\left(z_i^*\right)$
  - Aus Symmetriegründen $z_i^* = z_i$ möglich
- Problem: $F^{-1}(z)$ nicht immer analytisch ermittelbar

  Lösung durch andere <u>direkte Methoden</u>: Ersatzverteilung durch
  - Komposition aus ZV mit invertierbaren VF (LK 8.2.2)
    - Erlang-verteilte ZV: Summe (LK 8.2.3) von exponentiellen ZV
    - Hyperexponentiell verteilte ZV: Zufällige Wahl von exponentiellen ZV mit unterschiedlichem Parameter
  - invertierbare Approximation einer Verteilungsfunktion (z.B. rationale Funktion)
    - Abramowitz & Stegun: Handbook of Mathematical Functions

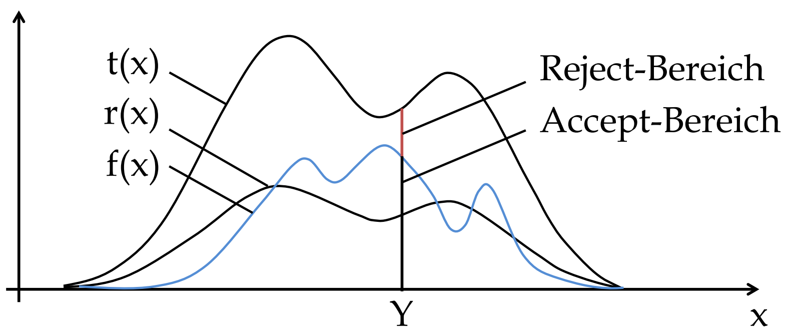
  Bei manchen Verteilungen nicht oder nicht effizient (Rechenzeit) anwendbar

3.1.4 Accept-Reject-Methode (LK 8.2.4)

- Indirekter Ansatz
  - Erzeuge eine Zufallszahl und generiere eine Stichprobe Y
  - Erzeuge eine weitere Zufallszahl und ermittle dadurch, ob Y als gesuchte Stichprobe X genommen werden kann
- Voraussetzung
  - Majorante $t(x)$ von $f(x)$, d.h. $\forall x\colon t(x) \geq f(x)$
    - Die Integralfunktion von t(x) soll leicht invertierbar sein
    - t(x) ist i.A. keine Dichtefunktion wegen

$$c = \int_{-\infty}^{\infty} t(x)\,dx \geq \int_{-\infty}^{\infty} f(x)\,dx = 1$$

  - Dichtefunktion r(x) durch Normierung: $r(x) = \frac{1}{c} \cdot t(x)$
  - Invertierbare Verteilungsfunktion: $R(x) = \int_{-\infty}^{x} r(t)dt$



- Algorithmus zur Erzeugung der ZV X
  - Erzeuge Y gemäß Verteilungsfunktion R(x)
  - Erzeuge $U \sim U(0,1)$, unabhängig von Y
  - Entscheide
    - X=Y falls $U \leq \frac{f(Y)}{t(Y)}$ (Accept)
    - Ansonsten verwerfe Y (Reject) und erneuter Versuch
- Y wird akzeptiert, falls Punkt $(Y, U \cdot t(Y))$ unter der Dichtekurve f(x) liegt.
- Wichtig für gute Effizienz: t(x) eng an f(x) anliegend

- Beweis der Gültigkeit des Algorithmus: LK Appendix 8A
  - X ~ f(x)
  - Majorante: t(x) ≥ f(x)
  - Y ~ $\mathrm{r(x)} = \frac{\mathrm{t(x)}}{\mathrm{c}} = \frac{\mathrm{t(x)}}{\int_{-\infty}^{\infty} \mathrm{t(x)dx}}$
  - Ereignis A: Y=y wird angenommen $\Leftrightarrow \mathrm{U} \le \frac{\mathrm{f(y)}}{\mathrm{t(y)}}$ mit U ~ U(0,1)

$$\Rightarrow \mathrm{P(A \mid Y = y)} = \frac{\mathrm{f(y)}}{\mathrm{t(y)}}$$

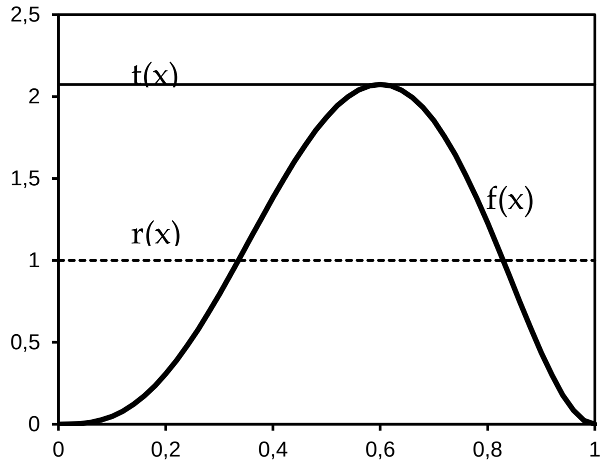
  - $$P(X \le x) = P(Y \le x \mid A) = \frac{P(A, Y \le x)}{P(A)} =$$

$$\frac{\int_{-\infty}^{x} P(Y \le x, A \mid Y = y) \cdot r(y) dy}{\int_{-\infty}^{\infty} P(A \mid Y = y) \cdot r(y) dy} =$$

$$= \frac{\int_{-\infty}^{x} P(A \mid Y = y) \cdot \frac{t(y)}{c} dy}{\int_{-\infty}^{\infty} \frac{f(y)}{t(y)} \cdot \frac{t(y)}{c} dy} = \frac{\int_{-\infty}^{x} \frac{f(y)}{t(y)} \cdot \frac{t(y)}{c} dy}{\int_{-\infty}^{\infty} \frac{f(y)}{c} dy} =$$

$$\frac{\frac{1}{c} \int_{-\infty}^{x} f(y) dy}{\frac{1}{c}} = \int_{-\infty}^{x} f(y) dy = F(x) \quad \text{q.e.d.}$$

- Beispiel: Erzeugung von ZV gemäß der beta(3,4) Verteilungsfunktion
- $f(x) = \begin{cases} 60x^3(1-x)^2 & 0 \le x \le 1 \\ 0 & \text{sonst} \end{cases}$



- Verteilungsfunktion ist Polynom höherer Ordnung und schwer zu invertieren.
- Maximum: (0.6 ; 2.0736)
- Majorante: $t(x) = \begin{cases} 2.0736 & 0 \le x \le 1 \\ 0 & \text{sonst} \end{cases}$

$$\Rightarrow c = \int_0^1 t(x)\,dx = 2.0736$$

$$\Rightarrow r(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{sonst} \end{cases} \sim U(0,1)$$

- Erzeugung der Zufallszahlen:
  - $Y \sim U(0,1)$
  - $U \sim U(0,1)$
  - $U \le \dfrac{60Y^3(1-Y)^2}{2.0736}$?

3.2 Kontinuierliche Verteilungsfunktionen (LK 6.2.2, 6.2.4 und 8.3)

3.2.1 Gleichverteilung: ZV X ~ U(a,b) (LK 8.3.1)

- VDF: $f(x)=1/(b-a)$, $X \in [a; b]$
- Range [a, b]
- VF: $F(x)=(x-a)/(b-a)$
- $E[X]=(a+b)/2$
- $VAR[X]=(b-a)^2/12$
- Mode: --
- Erzeugung durch Inversion: $U \sim U(0,1),\ X = a + (b-a)U$.

3.2.2 Exponentialverteilung: ZV X ~ expo(λ) (LK 8.3.2)

- VDF: $f(x) = \lambda \cdot e^{-\lambda \cdot x}$ für $x \geq 0$
- VF: $F(x) = 1 - e^{-\lambda \cdot x}$
- Range: $[0, \infty)$
- $E[X]=1/\lambda$
- $VAR[X]=1/\lambda^2$
- $c_{var}=1$
- Mode: 0
- Erzeugung durch Inversion: $U \sim U(0,1)$, $X= -\ln(U)/\lambda$
- Exponentiell verteilte Zwischenankunftszeiten ergeben einen **Poisson-Prozess**

3.2.2.1 Definition: Rate

- Wahrscheinlichkeit, dass ein Ereignis innerhalb eines Intervalls dt eintritt: $P(A \leq dt)$
- Rate=Wahrscheinlichkeit/Intervalllänge dt, wenn die Intervalllänge dt infinitesimal klein wird: $\lim_{dt \to 0} \frac{P(A \leq dt)}{dt} = \lambda$
- Die Rate der exponentiellen Verteilungsfunktion erfüllt diese Definition:
  $\lim_{dt \to 0} \frac{1-e^{-\lambda \cdot dt}}{dt} = \lim_{dt \to 0} \frac{1-\left(1-\frac{\lambda \cdot dt}{1!}+\frac{(\lambda \cdot dt)^2}{2!}-\cdots\right)}{dt} = \lambda$
  (hint: http://de.wikipedia.org/wiki/Taylorreihe)

3.2.2.2 Anwendung: Abstand von zufälligen (Beobachtungs)Punkten

Problem: Punkte sollen zufällig mit einem mittleren Abstand Δ und mit gleicher Dichte über eine beliebig lange Strecke verteilt sein. Welcher Verteilungsfunktion folgt der Abstand zwischen den Punkten?

Lösung

- Gegeben sei ein Kreis der Länge $n \cdot \Delta$ mit n zufälligen Punkten P_i, welche zufällig auf dem Kreis verteilt sind.

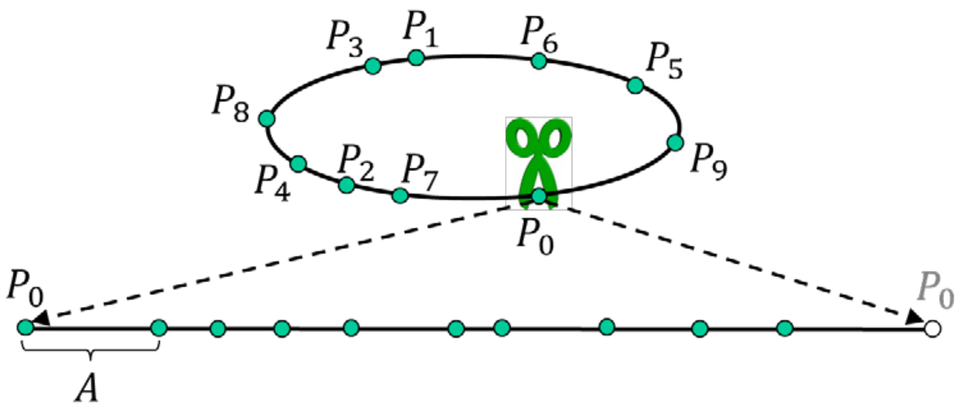


Abbildung 1: Der Abstand zwischen zufälligen Punkten auf einem Kreis ist gemäß Konstruktion identisch verteilt und entspricht der Zufallsvariablen *A*.

- Wir schneiden den Kreis o.B.d.A. an Punkt P_0 auf.
- Die Zufallsvariable D_i (für $0 < i < n$) bezeichne die Entfernung von P_i zu P_0 im Uhrzeigersinn.
- Gemäß Konstruktion ist sie zufällig und uniform zwischen 0 und $n \cdot \Delta$ verteilt und kann somit mit $D_i(t) = \frac{t}{n \cdot \Delta}$ angegeben werden.
- Die Zufallsvariable A bezeichne die Entfernung zwischen Punkt P_0 und seinem Nachbarn im Uhrzeigersinn.
- Die Verteilung von $A = \min\limits_{0<i<n} (D_i)$ berechnet sich berechnet als $A(t) = 1 - \prod_{0<i<n}\left(1 - D_i(t)\right) = 1 - \prod_{0<i<n}\left(1 - \frac{t}{n \cdot \Delta}\right) = 1 - \left(1 + \frac{-t}{n \cdot \Delta}\right)^{n-1}$.
- Für $n \to \infty$ erhalten wir $A(t) = \lim\limits_{n \to \infty} 1 - \frac{\left(1+\frac{-t}{n \cdot \Delta}\right)^n}{\left(1+\frac{-t}{n \cdot \Delta}\right)} = 1 - e^{\left(-\frac{1}{\Delta} \cdot t\right)}$ für ein festes t.
- Die Entfernung zwischen zufällig verteilten Punkten mit durchschnittlichem Abstand Δ folgt einer exponentiell Verteilungsfunktion mit Rate $\frac{1}{\Delta}$.

Folgerung 1: Zufällig und mit gleicher Dichte über eine beliebig lange Strecke verteilte Punkte können durch einen Poisson-Prozess realisiert werden.
Besonderheit: Gedächtnislosigkeit (Vorwärtsrekurrenzzeit ~ expo(λ))

Folgerung 2: Bei einem Poisson-Prozess ist zu jedem Zeitpunkt die Dauer bis zum nächsten Ereignis unabhängig von der Zeit seit dem letzten Ereignis.

3.2.3 Erlang-Verteilung: ZV X ~ k-Erlang(λ) (LK 8.3.3)

- ZV ist Summe von k expo(λ) verteilten ZVs (Phasen)

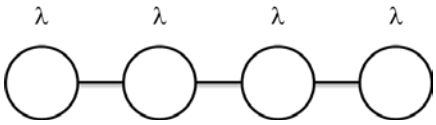


Abbildung 3-2: Phasendiagramm einer Erlang-4-Verteilung als Serienschaltung von 4 exponentiell verteilten Phasen.

- VDF: $f(x) = \begin{cases} \frac{(\lambda x)^{k-1}}{(k-1)!} \cdot \lambda \cdot e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$
- VF: $F(x) = \begin{cases} 1 - e^{-\lambda x} \cdot \sum\limits_{0 \leq i < k} \frac{(\lambda x)^i}{i!} & x \geq 0 \\ 0 & x < 0 \end{cases}$
- Range: $[0, \infty)$
- $E[X] = k/\lambda$
- $VAR[X] = k/\lambda^2$
- $c_{var}[X] = 1/\sqrt{k}$
- Mode: $\frac{k-1}{\lambda}$
- Erzeugung durch Inversion:

  $U_i \sim U(0,1)$, $X = \sum_{0 \leq i < k} -\frac{1}{\lambda} \cdot \ln(U_i) = -\frac{1}{\lambda} \cdot \ln\left(\prod_{0 \leq i < k} U_i\right)$

3.2.4 Hyperexponentialverteilung: ZV X ~ H(λ_i, p_i, $0 \le i < k$)

- ZV ist Kombination von k expo(λ) verteilten ZVs (Phasen)

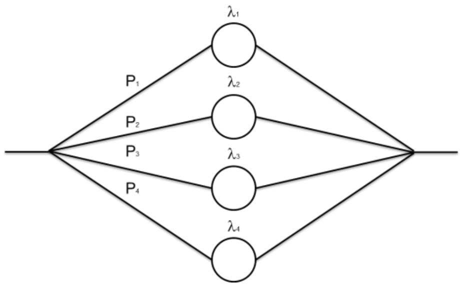


Abbildung 3-3: Phasendiagramm einer Hypoexponentiellen Verteilung: Parallelschaltung von 4 exponentiell verteilten Phasen, von denen genau eine ausgeführt wird. Vorgegebene Wahrscheinlichkeiten beeinflussen die zufällige Auswahl dieser Phase.

- VDF: $f(x) = \sum_{0 \le i < k} p_i \cdot \lambda_i \, e^{-\lambda_i \cdot x}$
- VF: $F(x) = 1 - \sum_{0 \le i < k} p_i \, e^{-\lambda_i \cdot x}$
- Range: $[0, \infty)$
- $E[X] = \sum_{0 \le i < k} (p_i / \lambda_i)$
- $c_{var}[X] = \sqrt{\frac{2 \cdot \sum_{0 \le i < k} (p_i / \lambda_i^2)}{(\sum_{0 \le i < k} (p_i / \lambda_i))^2} - 1}$ (aus Tran-Gia, Kapitel 1.3)
- Erzeugung durch $X = \begin{cases} X_0 \sim \text{expo}(\lambda_0) & \text{with probabilty } p_0 \\ \dots & \dots \\ X_{k-1} \sim \text{expo}(\lambda_{k-1}) & \text{with probabilty } p_{k-1} \end{cases}$ mit $\sum_{0 \le i < k} p_i = 1$
    - Erzeuge Zufallszahlen U_1, U_2
    - $i = min\left(m: \sum_{0 \le j \le m} p_j \ge U\right)$ // bestimme den Phasenindex
    - Benutze U_2 zur Erzeugung von X ~ expo(λ_i)
- Spezialfall: Hyperexponentielle Verteilung 2. Ordnung (H_2) zur Generierung einer ZV mit vorgegebenem E[X] und c_{var}[X]
    - Symmetriebedingung: $\frac{p1}{\lambda 1} = \frac{p2}{\lambda 2}$

      Geeignete Parameterwahl: $\lambda_{1,2} = \frac{1}{E[X]} \cdot \left(1 \pm \sqrt{\frac{c_{var}^2 - 1}{c_{var}^2 + 1}}\right)$

3.2.5 Normalverteilung: ZV X ~ N(μ, σ²) (LK 8.3.6)

- $f(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\left(\frac{(x-\mu)^2}{2 \cdot \sigma^2}\right)}$
- Range: (-∝, ∝)
- E[X]=μ, VAR[X]= σ²
- Skalierbarkeit: $X \sim \mathrm{N}(0,1) \Rightarrow (\mu + \sigma X) \sim \mathrm{N}(\mu, \sigma^2)$.
- Beispiel

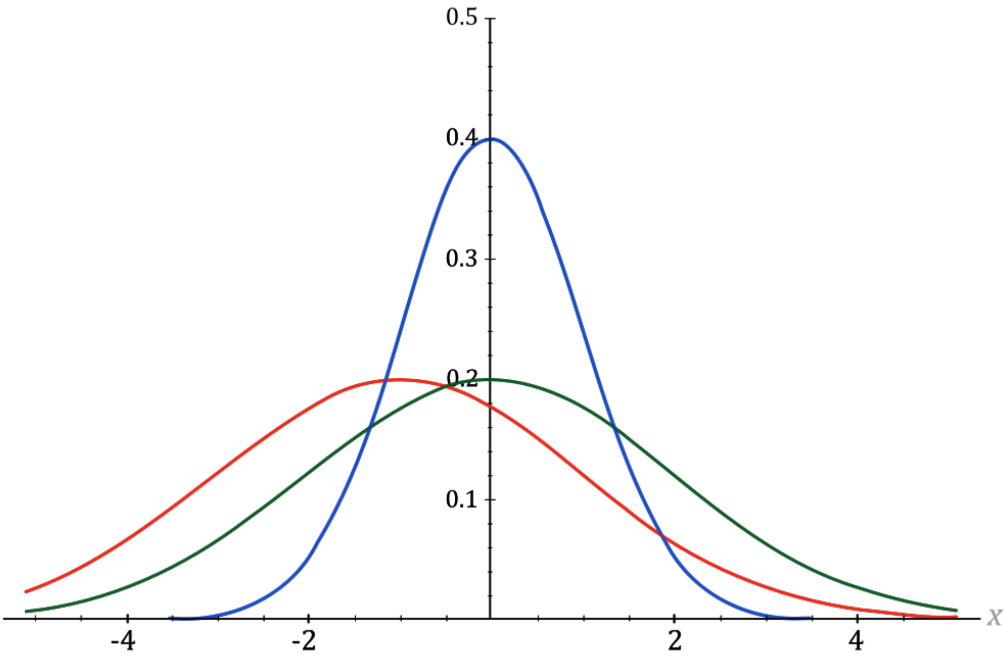


Abbildung 4: Dichtefunktionen der Normalverteilungen N(0,1), N(0,2), N(-1,2).

- Erzeugung durch Accept-Reject Methode

  $U_1, U_2 \sim U(0,1)$, wobei U_1 und U_2 unabhängig sein müssen.

  $V_i = 2U_i - 1, W = V_1^2 + V_2^2$.

  falls $W \leq 1$

  dann Accept: $Y = \sqrt{\frac{-2 \ln W}{W}}$, $X_1 = V_1 \cdot Y$, $X_2 = V_2 \cdot Y$ (2 Zufallsvariablen)

  sonst Reject

3.3 Diskrete Verteilungen (LK 6.2.3 und 8.4)

3.3.1 Deterministische Verteilung mit Parameter γ

- Verteilung: $x(i) = \begin{cases} 1 & f\ddot{u}r\ i = \gamma \\ 0 & sonst \end{cases}$
- E[X]= γ
- VAR[X]=0
- Range: γ
- Mode: γ

3.3.2 Diskrete Gleichverteilung: ZV ~ DU(i,j) (LK 8.4.2)

- Verteilung: $p(k) = \begin{cases} \frac{1}{j-i+1} & if\ k \in \{i, i+1, \ldots, j\} \\ 0 & otherwise \end{cases}$
- Range: $i \le k \le j$
- E[X]=(i+j)/2
- $VAR[X] = \frac{(j-i+1)^2 - 1}{12}$
- Mode: --
- Erzeugung durch Inversion:

  $U \sim U(0,\ 1)$

  $X = i + \lfloor (j-i+1) \cdot U \rfloor$

3.3.3 Beliebige diskrete Verteilung (LK 8.4.3)

- $p(x) = \begin{cases} p_k & if\ x = x_k, 0 \le k < n \\ 0 & otherwise \end{cases}$
- Range: beliebig aber diskret
  - Erzeugung durch Inversion:

    $U \sim U(0,1)$.

    $X = x_k$ mit $k = min(i{:}\sum_{0 \le j \le i} p_j \ge U)$

    Hoher Suchaufwand durch intelligente Datenstruktur vermeidbar

3.3.4 Bernoulli-Verteilung: ZV ~ Bernoulli(p) (LK 8.4.1)

- „Einfacher Münzwurf“
- Verteilung: $p(k)=\begin{cases}1-p & if \quad k=0\\ p & if \quad k=1\\ 0 & else\end{cases}$
- E[X]=p
- VAR[X]=p · (1-p)
- $c_{\text{var}}=\sqrt{\dfrac{1-p}{p}}$
- Mode: 0 bzw. 1
- Erzeugung durch Inversion:

  $\mathrm{U} \sim \mathrm{U}(0,\ 1)$

  Falls $U < p$, dann $X = 1$, sonst $X = 0$.

3.3.5 Binomial-Verteilung: ZV ~ Binom(n,p) (LK 8.4.4)

- $p(k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}$ mit $\binom{n}{k}=\dfrac{n!}{k!\cdot(n-k)!}$ und $0 \leq k \leq n$, $k \in N0$
- E[X]=n ·p
- VAR[X]=n ·p ·(1-p)
- $c_{\text{var}}[X]=\sqrt{\dfrac{1-p}{n\cdot p}}$
- Besonderheiten:
  - Für n→∝ ist Binom(n,p) durch N(E[X],VAR[X]) approximierbar
  - Für p<<1 und n sehr groß ist Binom(n,p) durch Poisson(E[X]) approximierbar
- Erzeugung durch Komposition wegen Binom(n, p) ~ $\sum_{0\leq i<n} Bernoulli(p)$:

  Erzeuge Y_i ~ Bernoulli(p), $X = \sum_{0\leq i<n} Y_i$

- Visualisierung

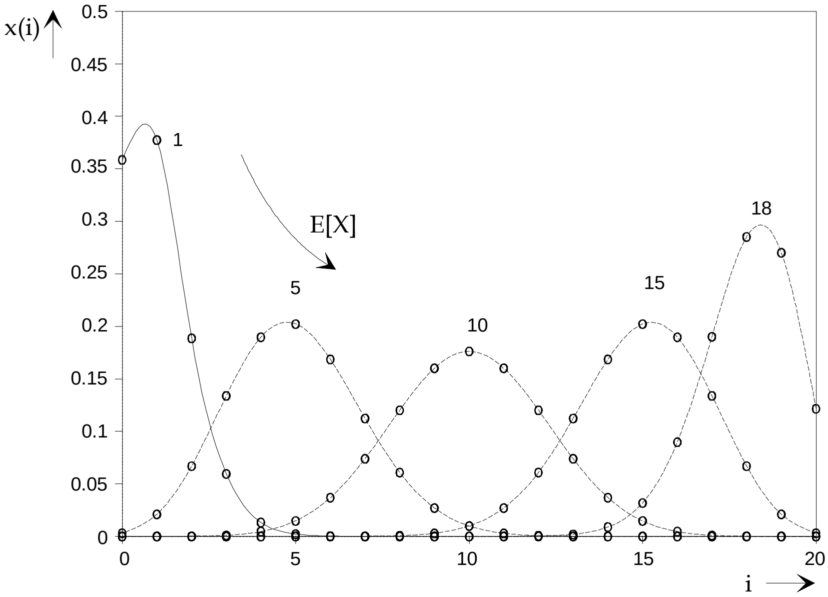


Abbildung 5: Binomialverteilungen.

3.3.6 Geometrische Verteilung: ZV ~ Geom(p) (LK 8.4.5)

- „Anzahl der Fehlversuche bis zum ersten Erfolg eines Bernoulli-Experiments mit Erfolgswahrscheinlichkeit p“
- $p(k) = (1-p)^k \cdot p$, k ≥ 0
- VF: $F(x) = 1-(1-p)^{\lfloor x \rfloor + 1}$

  Beweis: Geometrische Reihe: $\sum_{i=0}^{k} q^i = \frac{1-q^{k+1}}{1-q}, \sum_{i=0}^{\infty} q^i = \frac{1}{1-q}$; setze q=1-p;

- $$F(x) = P(X \le (k = \lfloor x \rfloor)) = p \cdot \sum_{i=0}^{k} (1-p)^i = p \cdot \frac{1-(1-p)^{k+1}}{1-(1-p)}$$
  $$= p \cdot \frac{1-(1-p)^{k+1}}{p} = 1-(1-p)^{\lfloor x \rfloor + 1}$$
- E[X]=(1-p)/p
- VAR[X]=(1-p)/p²
- $c_{\text{var}}[X] = \frac{1}{\sqrt{(1-p)}}$
- Mode 0

- Erzeugung durch Inversion:

  U ~ U(0,1)

  $$X = \left\lfloor \frac{\ln(U)}{\ln(1-p)} \right\rfloor$$

- Visualisierung

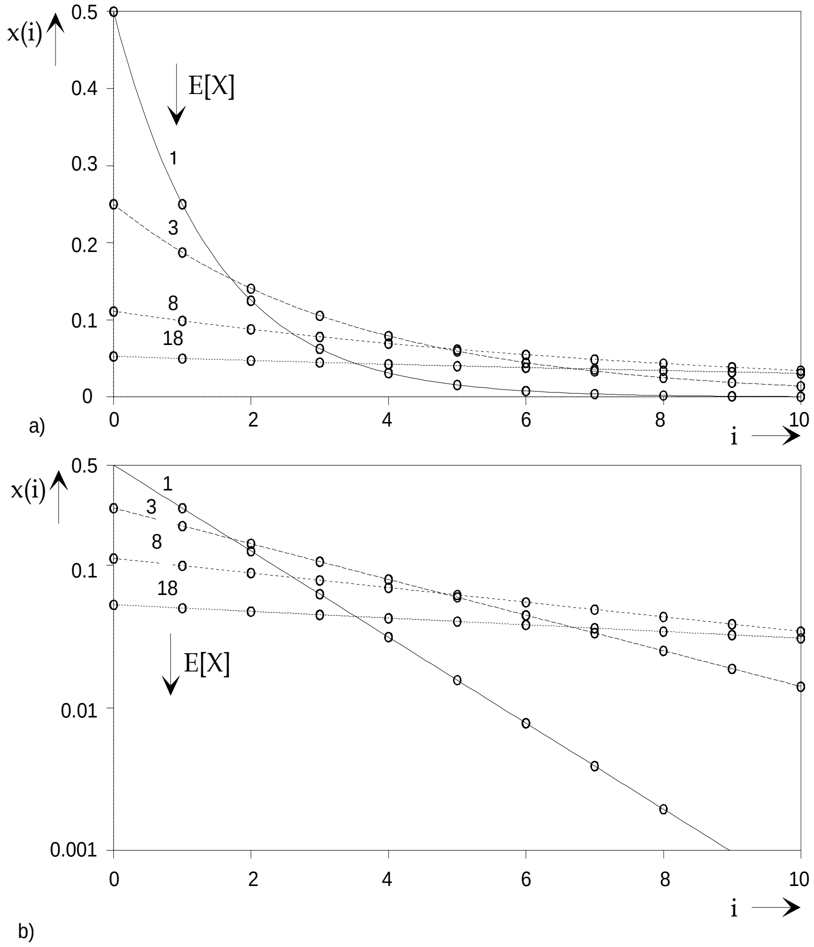


Abbildung 6: Geometrische Verteilungen.

3.3.7 „Um 1 Verschobene" Geometrische Verteilung

- „Anzahl aller Versuche bis das Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p einen Erfolg zeigt"
- $p(k) = (1-p)^{k-1} \cdot p$, k ≥ 1
- E[X]=(1-p)/p + 1=1/p
- VAR[X]=(1-p)/p² (dieselbe wie bei der nicht verschobenen Geom)
- $c_{var}[X] = \sqrt{1-p}$

3.3.8 Negativ-binomiale Verteilung: ZV ~ NegBin(s,p) (LK 8.4.6)

http://en.wikipedia.org/wiki/Negative_binomial_distribution

- „Anzahl der Fehlversuche bis sich s Erfolge bei Bernoulli-Experimenten eingestellt haben. Die s erfolgreichen gehören wie bei der Geom(p)-Verteilung nicht dazu!“
- NegBin(s,p) ~ $\sum_{0 \le i < s} \mathrm{Geom}(p)$
- $p(k) = \binom{s+k-1}{k} \cdot p^s \cdot (1-p)^k = \binom{-s}{k} \cdot p^s \cdot (-(1-p))^k$

  Zur Umformung : http://de.wikipedia.org/wiki/Binomialkoeffizient
- E[X]=s·(1-p)/p
- VAR[X]=s· (1-p)/p²
- $c_{var}[X] = \frac{1}{\sqrt{s \cdot (1-p)}}$
- Mode: $\begin{cases} y \text{ und } y+1 & \text{falls } y \text{ ganzzahlig} \\ \lfloor y \rfloor + 1 & \text{sonst} \end{cases}$ mit y=[s*(1-p)-1]/p,
- Erzeugung durch Komposition:

  $Y_1, Y_2, \ldots, Y_s \sim \mathrm{geom}(p)$ und unabhängig.

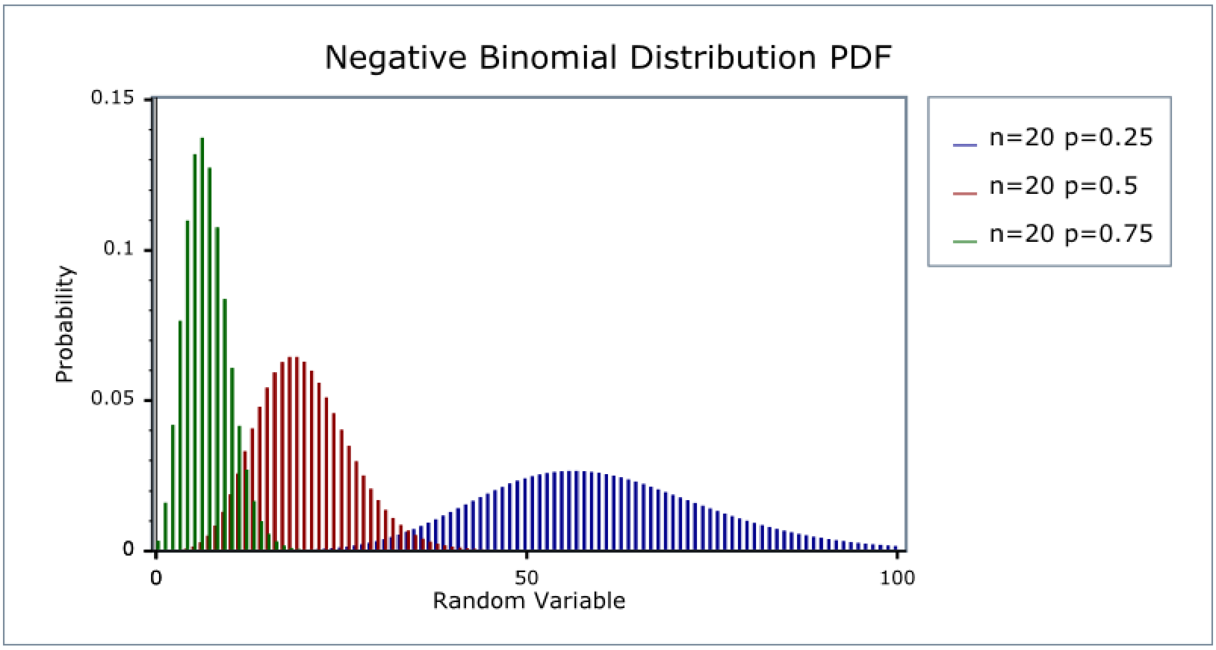
  $X = Y_1 + Y_2 + \cdots + Y_s$.
- Erweiterung auf reelle s möglich:

  $p(k) = \frac{\Gamma(s+k)}{\Gamma(s) \cdot k!} \cdot p^s \cdot (1-p)^k$ im Spezialfall s ganzzahlig: Γ(s)=(s-1)! (siehe später!)
- Der Grenzwert der Negbin für sehr große p und s nähert sich einer Poisson-Verteilung an:

  Substituiere $p = \frac{\omega}{\lambda + \omega}$ und $s = \omega$

  Dann gilt: $p(k) = \frac{y^k}{k!} \cdot \frac{\Gamma(\omega + k)}{\Gamma(\omega) \cdot (y+\omega)^k} \cdot \frac{1}{\left(1 + \frac{y}{\omega}\right)^{\omega}}$

  und im Grenzfall bekommen wir $\lim_{\omega \to \infty} p(k) = \frac{y^k}{k!} \cdot 1 \cdot \exp(-y)$ (Poissonverteilung)

- 

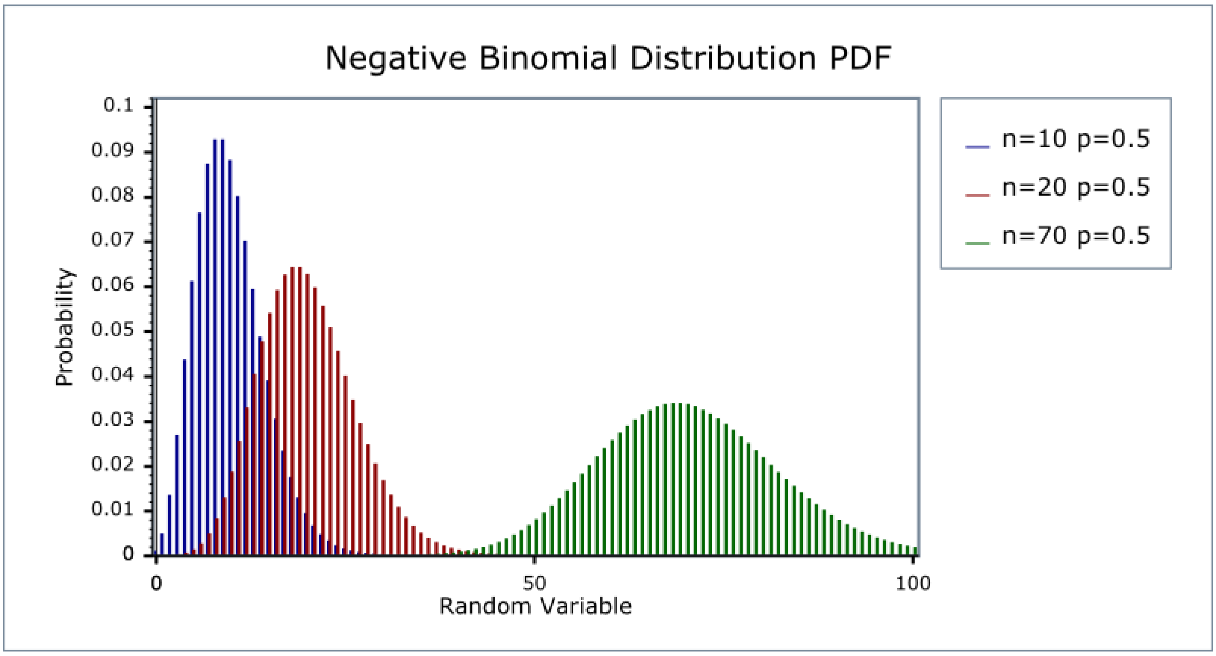


Abbildung 7: Negativ-binomiale Verteilung: Einfluss der Parameter.

Quelle: http://www.boost.org/doc/libs/1_51_0/libs/math/doc/sf_and_dist/html/math_toolkit/dist/dist_ref/dists/negative_binomial_dist.html

3.3.9 Poisson-Verteilung: ZV ~ Poisson(y) (LK 8.4.7)

- Anzahl der Ankünfte eines Poisson-Prozesses (= exponentielle Zwischenankunftszeiten) mit Rate λ in einem Intervall der Länge τ und y=λ*τ
- $p(k) = \frac{y^k}{k!} \cdot e^{-y}$

  Mit $e^y = \sum_{0 \le k < \infty} \frac{y^k}{k!}$ ist klar, dass es sich um eine Verteilung handelt.
- E[X]= y

  Beweis: $E[X] = \sum_{0 \le k < \infty} k \cdot \frac{y^k}{k!} \cdot e^{-y} = y \cdot e^{-y} \cdot \sum_{1 \le k < \infty} \frac{y^{k-1}}{(k-1)!} = y \cdot e^{-y} \cdot e^y = y$
- VAR[X]= y
- $c_{\text{var}}[X] = \frac{1}{\sqrt{y}}$
- Mode $\begin{cases} y \text{ und } y-1 & \text{falls } y \text{ ganzzahlig} \\ \lfloor y \rfloor & \text{sonst} \end{cases}$
- Beispiel

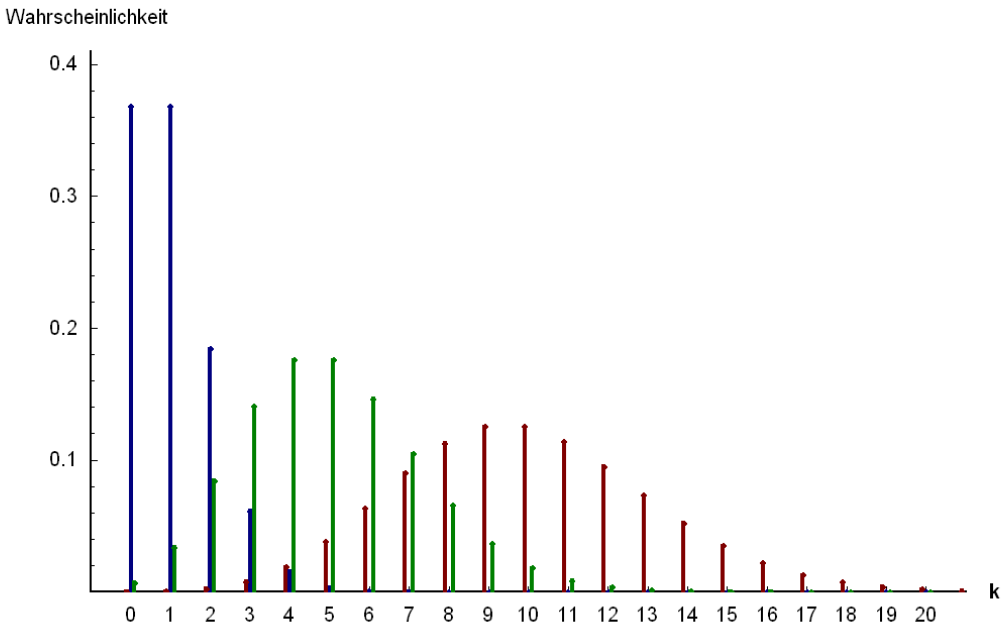


Abbildung 8: Wahrscheinlichkeitsfunktion der Poisson-Verteilung für λ = 1 (blau), λ = 5 (grün) und λ = 10 (rot); Quelle: http://de.wikipedia.org/wiki/Poisson-Verteilung

- Erzeugung durch Quasi-Komposition: Idee τ=1 ⇒ λ=y

  $Y_j \sim \text{expo}(y) \Rightarrow X = \max\left\{k : \sum_{j=1}^{k} Y_j \le 1\right\} \sim \text{Poisson}(y).$

  $$\begin{array}{ccccc} \sum\limits_{0\le j<k} Y_j & \le & 1 & < & \sum\limits_{0\le j\le k} Y_j \\ \sum\limits_{0\le j<k}\left(-\frac{1}{y}\ln(U_j)\right) & \le & 1 & < & \sum\limits_{0\le j\le k}\left(-\frac{1}{y}\ln(U_j)\right) \\ \ln\left(\prod\limits_{0\le j<k} U_j\right) & \ge & -y & > & \ln\left(\prod\limits_{0\le j\le k} U_j\right) \\ \prod\limits_{0\le j<k} U_j & \ge & e^{-y} & > & \prod\limits_{0\le j\le k} U_j \end{array}$$

  Algorithmus:

  - $a = e^{-y}, b = 1, k = -1.$
  - Wiederhole
  - $k = k+1, \quad U_i \sim U(0,1), \quad b = b \cdot U_k$
  - bis $b \le a$
  - X=k

- Achtung: e^{-y} kann sehr klein werden. Beispiel: y=1000 ⇒ $e^{-y}=5\cdot 10^{-435}$

  Dann ergeben sich numerische Fehler, die zu falschen Mittelwerten und Varianzen führen. Abhilfe durch clevere Implementierung, die große Zahlen vermeidet.

  - $$\begin{array}{ccc} \prod\limits_{0\le j<k} U_j & \ge e^{-y} > & \prod\limits_{0\le j\le k} U_j \\ e^y \cdot \prod\limits_{0\le j<k} U_j & \ge 1 > & e^y \cdot \prod\limits_{0\le j\le k} U_j \end{array}$$
  - Idee
    - $e^y = \left(e^{10}\right)^{(y \operatorname{div} 10)} \cdot e^{(y \bmod 10)}$
    - Multipliziere abwechselnd Zufallszahlen U_j und (y div 10) Mal e^{10} bzw. $e^{(y \bmod 10)}$, so dass das Produkt nie kleiner als eine gewisse Schranke wird.
  - Implementierung siehe Übung!

3.4 Verwandtschaft kontinuierlicher und diskreter Verteilungen

Abbildung 9: Zeitkontinuierlicher Ankunftsprozess.

Wir nehmen an, die Zeit ist kontinuierlich und zu jedem Zeitpunkt ist mit derselben Wahrscheinlichkeitsdichte eine Ankunft möglich. Dies wird in Abbildung 9 illustriert.

- Die Zeit zwischen Ankünften ist exponentiell verteilt mit Rate λ (Expo(λ), durchschnittliche Zwischenankunftszeit ist $1/\lambda$).
- Die Zeit, bis sich k Ankünfte ereignen, ist Erklang-k verteilt (Erlang(k, λ)).
- Die Anzahl der Ankünfte innerhalb eines Intervalls fester Länge τ ist Poisson-verteilt mit Parameter $y=\tau*\lambda$ (Poisson(y)).

Abbildung 10: Zeitdiskreter Ankunftsprozess, bei dem rote Slots Ankünfte darstellen.

Wir nehmen an, die Zeit ist in Slots eingeteilt und in jedem Slot ist mit Wahrscheinlichkeit p eine Ankunft möglich wie Abbildung 10 illustriert.

- Die Anzahl der Slots zwischen Ankünften ist geometrisch Geom(p) verteilt mit $p=1/(E[X]+1)$.
- Die Anzahl der freien Slots, bis s Ankünfte geschehen sind, ist negativ-binomial verteilt (NegBin(s,p)).
- Die Anzahl der Ankünfte innerhalb eines Intervalls fester Länge n ist binomial verteilt (Binom(n,p)).

Somit begründen wir die Verwandtschaft der entsprechenden Verteilungen.

3.5 Übersicht behandelter Verteilungen

3.5.1 Kontinuierliche Verteilungen

1. Gleichverteilung
2. Exponentielle Verteilung (λ) (cvar=1)
   - „zufällige" Zwischenankunftszeiten (Anrufe)
3. Erlang-k-Verteilung (k, λ) (cvar<1)
   - Summe von k exponentiellen Phasen
4. Hyperexponentielle Verteilung (cvar>1)
5. Normalverteilung (μ,σ)
   - Summe sehr vieler iid Zufallsvariablen

3.5.2 Diskrete Verteilungen

1. Diskrete Gleichverteilung
2. Beliebige diskrete Verteilung
3. Bernoulli-Verteilung Bernoulli(p)
   - Münzwurf
4. Binomial-Verteilung Binom(n,p)
   - Wiederholter Münzwurf
   - Ziehen von Kugeln mit Zurücklegen
   - Ankünfte mit geometrisch verteilten Zwischenankunftszeiten Geom(p) in einem Intervall ganzzahliger Länge n
5. Geometrische Verteilung Geom(p)
   - Wiederholter Münzwurf bis 1 Mal gewonnen wurde
   - ganzzahliges Gegenstück zur exponentiellen Verteilung
6. Negativ-binomiale Verteilung NegBin(s,p)
   - Wiederholter Münzwurf: Anzahl von Fehlversuchen bis s Mal gewonnen wurde
7. Poisson-Verteilung Poisson(y)
   - Anzahl von Ankünften mit exponentiell verteilten Zwischenankunftszeiten in einem Intervall fester Länge