

Ue Statistik u. Wahrscheinlichkeitsth. f. Inf. 107.251 W 2005 6 http://www.statistik.tuwien.ac.at/RISueb/	Di 12:00 – 18:00
	9.Blatt
Werner GURKER Tel.: 58801-10724 Spr.: Di u. Do 11-12 e-mail: W.Gurker@tuwien.ac.at	10. Jän. 2006

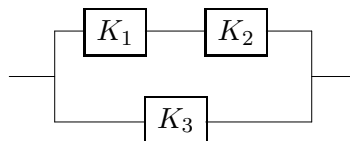
- Der Input Λ eines Programms sei eine gammaverteilte sG mit den Parametern α und β , $\Lambda \sim \text{Gam}(\alpha, \beta)$. Bedingt durch $\Lambda = \lambda$ ist die Ausführungszeit des Programms eine gammaverteilte sG mit den Parametern k und $1/\lambda$:

$$X|\Lambda = \lambda \sim \text{Gam}(k, 1/\lambda)$$

Ermitteln Sie die Dichte der Ausführungszeit X des Programms. Was ergibt sich speziell für $k = 1$, d.h. für $X|\Lambda = \lambda \sim \text{Ex}_{1/\lambda}$?

Hinweis: Bestimmen Sie zuerst die gemeinsame Dichte $f(x, \lambda)$ von (X, Λ) ; führen Sie das zu berechnende Integral $f_X(x) = \int_0^\infty f(x, \lambda) d\lambda$ auf die Gammafunktion zurück.

- Die logische Struktur eines Systems bestehend aus drei Komponenten sei gegeben wie folgt:



Die Lebensdauern der Komponenten seien unabhängig und identisch verteilt mit Dichte $f(x) = e^{-x} I_{(0, \infty)}(x)$. Bestimmen Sie die Dichte der Lebensdauer des Systems und deren Mittelwert.

Hinweis: Ermitteln Sie zuerst die Verteilungsfunktion.

- Die Lebensdauer X einer Komponente folge einer uniformen Verteilung auf dem Intervall $(0, 3)$. Fällt die Komponente aus, wird sie sofort durch eine Reservekomponente mit auf dem Intervall $(0, 5)$ uniform verteilter Lebensdauer Y ersetzt.

- Bestimmen (und zeichnen) Sie die Dichte der Gesamtlebensdauer, d.h. bestimmen Sie die Dichte von $S = X + Y$.

Hinweis: Faltung; orientieren Sie sich am Beispiel der Vorlesung.

- Ermitteln Sie Mittelwert und Streuung von S .

- Ein Programm bestehe aus $n = 100$ Seiten Code und X_i sei die Anzahl der Fehler auf der i -ten Seite. Angenommen, die X_i 's sind unabhängig und identisch poissonverteilt mit Mittel $\mu = 0.8$. Wenn $Y = \sum_{i=1}^n X_i$ die Gesamtzahl der Fehler ist, bestimmen Sie mit Hilfe des Zentralen Grenzwertsatzes einen approximativen Wert für $W\{75 < Y < 85\}$. Verwenden Sie bei der Berechnung die Stetigkeitskorrektur.

- Ein Hobbyphotograph findet heraus, daß (bei größtmöglicher Auflösung und geringster Kompression) seine Digitalbilder im Durchschnitt 1.92 MB groß sind, bei einer Streuung von 0.425 MB.

- (a) Mit welcher Wahrscheinlichkeit passen auf eine 128 MB Speicherkarte zumindest 61 Bilder (nach Herstellerangaben unter den gegebenen Bedingungen der Standardwert)?
- (b) Wieviele Bilder passen mit einer Wahrscheinlichkeit von zumindest 0.5 auf die Speicherkarte? Wieviele mit einer Wahrscheinlichkeit von 0.95 ?

Hinweis: Zentraler Grenzwertungssatz.

6. Für die Generierung von (approximativ) nach $N(0, 1)$ verteilten Zufallszahlen gibt es auch die folgende Methode: Man generiert 12 uniform (auf $(0, 1)$) verteilte Zufallszahlen U_i und bildet:

$$\sum_{i=1}^{12} U_i - 6$$

Dies liefert eine $N(0, 1)$ -verteilte Zufallszahl; für die Generierung einer weiteren nimmt man wieder 12 (neue) uniform verteilte Zufallszahlen, bildet den obigen Ausdruck, usw.

- (a) Geben Sie eine Erklärung für diese Methode. (*Bem.:* Diese Methode kommt ohne die Invertierung der Verteilungsfunktion aus; allerdings braucht man für die Erzeugung *einer* normalverteilten Zahl 12 uniform verteilte Zahlen.)
- * (b) Erzeugen Sie nach obiger Methode 500 normalverteilte Zufallszahlen mit Mittelwert 2 und Streuung 0.4 und stellen Sie das Ergebnis in Form eines Dichtehistogramms dar (mit darüber gezeichneter theoretischer Dichte).

Beispiel(teil)e mit (*) werden im Konversatorium behandelt.

1. Laut Angabe gilt:

$$f_{\Lambda}(\lambda) = \frac{\lambda^{\alpha-1} e^{-\lambda/\beta}}{\Gamma(\alpha) \beta^{\alpha}}, \quad \lambda > 0$$

und

$$f(x|\Lambda = \lambda) = f(x|\lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}, \quad x > 0$$

Für die gemeinsame Dichte von (X, Λ) folgt daher:

$$f(x, \lambda) = f(x|\lambda) f_{\Lambda}(\lambda) = \frac{x^{k-1} \lambda^{\alpha+k-1} e^{-\lambda(1+\beta x)/\beta}}{\Gamma(\alpha) \Gamma(k) \beta^{\alpha}}, \quad x > 0, \lambda > 0$$

Die (Rand-) Dichte von X ergibt sich durch Integration nach λ :

$$f_X(x) = \int_0^{\infty} f(x, \lambda) d\lambda = \frac{x^{k-1}}{\Gamma(\alpha) \Gamma(k) \beta^{\alpha}} \int_0^{\infty} \lambda^{\alpha+k-1} e^{-\lambda(1+\beta x)/\beta} d\lambda$$

Dieses Integral lässt sich auf die Gammafunktion zurückführen:

$$\frac{\lambda(1+x\beta)}{\beta} = u \quad \longrightarrow \quad d\lambda = \frac{\beta}{1+x\beta} du$$

Damit folgt:

$$\begin{aligned} f_X(x) &= \frac{\beta^k x^{k-1}}{\Gamma(\alpha) \Gamma(k) (1+x\beta)^{\alpha+k}} \int_0^{\infty} u^{\alpha+k-1} e^{-u} du \\ &= \frac{\Gamma(\alpha+k) \beta^k x^{k-1}}{\Gamma(\alpha) \Gamma(k) (1+x\beta)^{\alpha+k}}, \quad x > 0 \end{aligned}$$

Für $k = 1$ ergibt sich:

$$f_X(x) = \frac{\alpha \beta}{(1+\beta x)^{\alpha+1}}, \quad x > 0$$

Bem.: Eine Verteilung mit der obigen allgemeinen Dichte nennt man *verallgemeinerte Pareto-Verteilung* (ist auch eine Verallgemeinerung der F -Verteilung); für $k = 1$ ergibt sich die schon bekannte (einfache) *Pareto-Verteilung* (vgl. **Bsp 7.5**). Diese Verteilungen haben schwerere Ausläufer als die ursprüngliche (bedingte) Gammaverteilung.

2. Bezeichnet X_i die Lebensdauer der i -ten Komponente und F_i die zugehörige Verteilungsfunktion, so gilt laut Angabe:

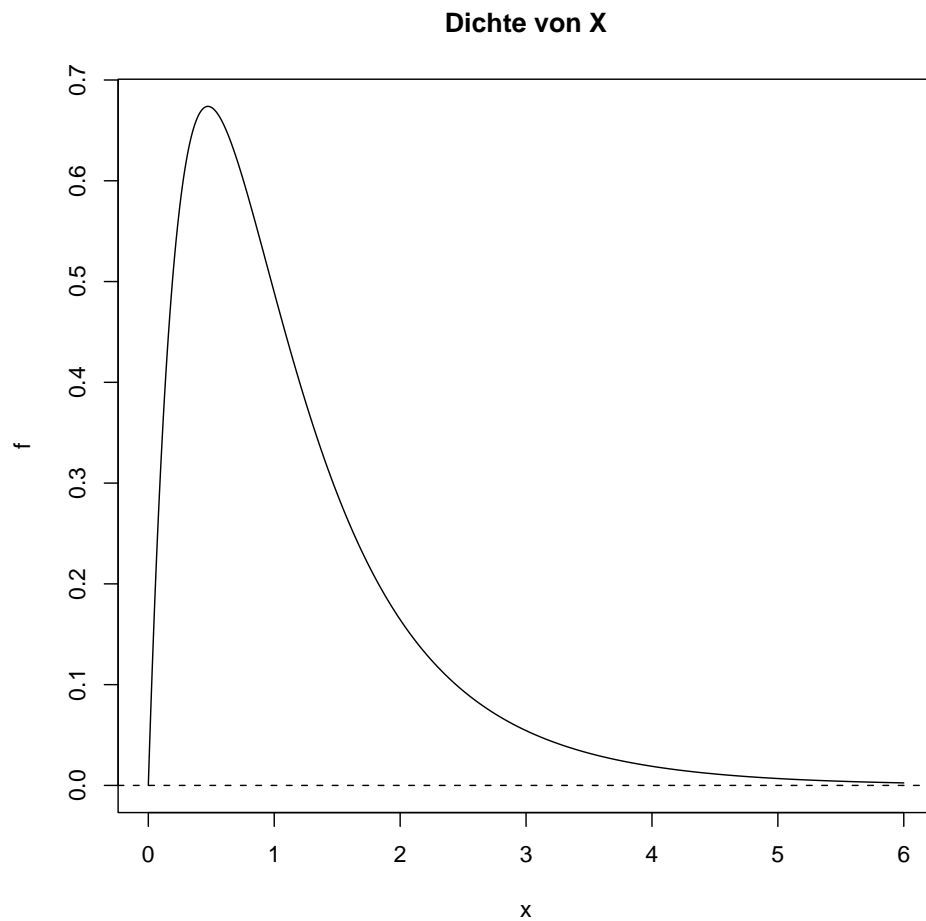
$$F_i(x) = 1 - e^{-x}, \quad x > 0$$

Damit folgt für die Verteilungsfunktion der Lebensdauer X des Systems:

$$\begin{aligned} F_X(x) &= W(X \leq x) = W(\max\{\min\{X_1, X_2\}, X_3\} \leq x) \\ &= [1 - (1 - F_1(x))(1 - F_2(x))] F_3(x) \\ &= (1 - e^{-2x})(1 - e^{-x}) \\ &= 1 - e^{-x} - e^{-2x} + e^{-3x}, \quad x > 0 \end{aligned}$$

Die Dichte bekommt man durch Ableiten:

$$f_X(x) = F'_X(x) = e^{-x} + 2e^{-2x} - 3e^{-3x}, \quad x > 0$$



Mittelwert von X :

$$\mathbb{E}(X) = \int_0^{\infty} x f_X(x) dx = \int_0^{\infty} x [e^{-x} + 2e^{-2x} - 3e^{-3x}] dx = 1 + \frac{1}{2} - \frac{1}{3} = \frac{7}{6}$$

Bem.: Bei der Berechnung des Mittelwerts wurde wiederholt verwendet, daß:

$$\int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \quad (= \text{Mittelwert einer } Ex_{1/\lambda}\text{-Verteilung})$$

3. (a) Die Dichte von $S = X + Y$ ergibt sich durch Faltung von f_X und f_Y :

$$f_X(x) = \frac{1}{3} I_{(0,3)}(x), \quad f_Y(y) = \frac{1}{5} I_{(0,5)}(y)$$

$$f_S(z) = \int_{-\infty}^{\infty} f_X(z-t) f_Y(t) dt = \frac{1}{15} \int_{-\infty}^{\infty} I_{(0,3)}(z-t) I_{(0,5)}(t) dt$$

Der Integrand ist nur dann ungleich Null, wenn $0 < z-t < 3$ und $0 < t < 5$. Die erste Bedingung ist äquivalent zu $z-3 < t < z$; beide Bedingungen sind also nur dann erfüllt, wenn:

$$\max\{0, z-3\} < t < \min\{5, z\}$$

Fallunterscheidung ($f_S(z) = 0$ für $z \leq 0$ und $z > 8$):

(1) $0 < z \leq 3$: $\max\{0, z-3\} = 0$, $\min\{5, z\} = z$

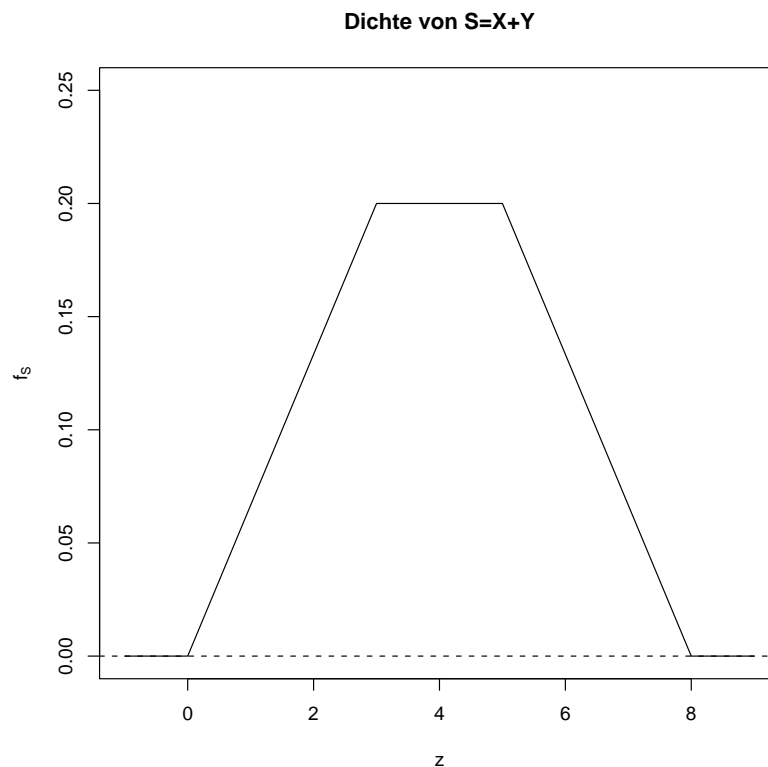
$$f_S(z) = \frac{1}{15} \int_0^z dt = \frac{z}{15}$$

(2) $3 < z \leq 5$: $\max\{0, z-3\} = z-3$, $\min\{5, z\} = z$

$$f_S(z) = \frac{1}{15} \int_{z-3}^z dt = \frac{1}{5}$$

(3) $5 < z \leq 8$: $\max\{0, z-3\} = z-3$, $\min\{5, z\} = 5$

$$f_S(z) = \frac{1}{15} \int_{z-3}^5 dt = \frac{8-z}{15}$$



- (b) Der Mittelwert einer Summe ist die Summe der Mittelwerte:

$$\mathbb{E}(S) = \mathbb{E}(X) + \mathbb{E}(Y) = 1.5 + 2.5 = 4$$

Wegen der Unabhängigkeit von X und Y gilt:

$$\text{Var}(S) = \text{Var}(X) + \text{Var}(Y) = \frac{9}{12} + \frac{25}{12} = \frac{34}{12}$$

Bem.: Für die Gültigkeit der obigen Beziehung genügt die Unkorreliertheit von X und Y .

Die Streuung ist die (positive) Wurzel aus der Varianz:

$$\text{Streuung}(S) = \sqrt{\frac{34}{12}} \doteq 1.683$$

4. Auf Basis des Zentralen Grenzwertungssatzes (ZGVS) gilt (± 0.5 ist die Stetigkeitskorrektur):

$$\begin{aligned} W\{75 < Y < 85\} &= W\{76 \leq Y \leq 84\} = \sum_{i=76}^{84} W\{Y = i\} \\ &\approx \Phi\left(\frac{84.5 - 100 \cdot 0.8}{\sqrt{100 \cdot 0.8}}\right) - \Phi\left(\frac{75.5 - 100 \cdot 0.8}{\sqrt{100 \cdot 0.8}}\right) \\ &= \Phi\left(\frac{84.5 - 80}{\sqrt{80}}\right) - \Phi\left(\frac{75.5 - 80}{\sqrt{80}}\right) \\ &= 2\Phi(0.5031) - 1 \\ &= 0.3851 \end{aligned}$$

Bem.: Nach dem Additionstheorem für (unabhängige) Poissonverteilungen gilt:

$$Y = \sum_{i=1}^{100} X_i \sim P_{100 \times 0.8} = P_{80}$$

Der exakte Wert für die gesuchte Wahrscheinlichkeit ist (berechnet mit R):

$$W\{76 \leq Y \leq 84\} = \sum_{i=76}^{84} \frac{80^i e^{-80}}{i!} = 0.3850$$

5. (a) Ist N die Anzahl der auf die Speicherkarte passenden Bilder, und sind X_i die Größen der einzelnen Bilder, so gilt nach dem ZGVS:

$$\begin{aligned} W\{N \geq 61\} &= W\left\{\sum_{i=1}^{61} X_i \leq 128\right\} \approx \Phi\left(\frac{128 - 61 \cdot 1.92}{\sqrt{61 \cdot 0.425}}\right) \\ &= \Phi(3.2777) \doteq 0.9995 \end{aligned}$$

(b) In diesem Fall ist das größte n zu bestimmen, sodaß:

$$W\{N \geq n\} = W\left\{\sum_{i=1}^n X_i \leq 128\right\} \approx \Phi\left(\frac{128 - n \cdot 1.92}{\sqrt{n} \cdot 0.425}\right) \geq 0.5$$

Man findet schnell heraus, daß:

$$W\{N \geq 67\} \doteq 0.4270 \quad W\{N \geq 66\} \doteq 0.6446$$

D.h., in ca. 64% der Fälle bringt der Fotograf (unter den gegebenen Bedingungen) zumindest 66 Bilder unter. Ebenso gilt:

$$W\{N \geq 64\} \doteq 0.9340 \quad W\{N \geq 63\} \doteq 0.9816$$

D.h., in ca. 98% der Fälle bringt er zumindest 63 Bilder unter.

6. (a) Die Begründung liegt im ZGVS; für den Erwartungswert und die Varianz einer nach $U_{(0,1)}$ verteilten sG U gilt:

$$\mathbb{E}(U) = \frac{1}{2}, \quad \text{Var}(U) = \frac{1}{12}$$

Für die Summe von 12 unabhängigen $U_{(0,1)}$ -verteilten Größen U_i folgt daher:

$$\mathbb{E}\left(\sum_{i=1}^{12} U_i\right) = 6, \quad \text{Var}\left(\sum_{i=1}^{12} U_i\right) = 1$$

Nach dem ZGVS gilt:

$$\sum_{i=1}^{12} U_i - 6 \sim N(0, 1)$$

Bem.: Auf Grund der Symmetrie der Ausgangsverteilung ($U_{(0,1)}$) ist die Normalapproximation bereits für den relativ kleinen Stichprobenumfang $n = 12$ sehr gut.

- *(b) Die folgende R-Funktion erzeugt normalverteilte Zufallszahlen nach der obigen Methode; soll der Mittelwert gleich μ und die Varianz gleich σ^2 sein, muß man diese Zahlen noch mit σ multiplizieren und μ addieren. (*Bem.:* R-Skript unter **b96.r** auf der Ü-Homepage.)

```
rand.norm <- function(n,mu,sigma) {
  r.n <- numeric(n)
  for (i in (1:n)) {
    r.u <- runif(12)
    r.n[i] <- sum(r.u)-6 }
  r.n <- mu+r.n*sigma }

mu <- 2; sigma <- 0.4; n <- 500
res <- rand.norm(n,mu,sigma)
tmp.hist <- hist(res,plot=FALSE)
tmp.dens <- dnorm(mu,mean=mu,sd=sigma)
y.max <- max(tmp.hist$density,tmp.dens)
hist(res,breaks=seq(mu-4*sigma-1,mu+4*sigma+1,length=round(sqrt(n))+1),
     prob=TRUE,xlab="",ylab="",main=paste("Mittel=",mu,"  Streuung=",
     sigma,"  n=",n),ylim=c(0,y.max))
x <- seq(mu-4*sigma,mu+4*sigma,length=100)
lines(x,dnorm(x,mu,sigma),lty=1,lwd=2,col="red")
rug(res)
```

Die Abbildung zeigt ein (Dichte-) Histogramm von 500 so erzeugten Zufallszahlen einer $N(2, 0.16)$ -Verteilung (mit darüber gezeichneter Normaldichte und einem „Rugplot“):

