

2. Teilprüfung

Statistik und Wahrscheinlichkeitsrechnung

FS 2008

Dr. J. Köhler

ETH Zürich

Dienstag 20. Mai 2008

08:15 – 09:45

Lösungen

Teil 1: Multiple Choice (maximal 31 Punkte)

In den folgenden Multiple Choice Fragen können für die gleiche Frage mindestens eine oder mehrere Antworten richtig sein.

- 1.1** Im Kanton Zug gibt es pro Jahr im Durchschnitt drei Grossbrandereignisse, welche zu Sachschäden grösser als 1 Mio. CHF führen. Das Auftreten dieser Brände kann mit einem Poissonprozess beschrieben werden. Die Anzahl n der Grossbrandereignisse in t Jahren wird durch die kumulative Verteilungsfunktion $P(N = n) = \frac{(\nu t)^n}{n!} e^{-\nu t}$ mit der mittleren jährlichen Rate ν beschrieben.

(2 Punkte)

Der Prozess unterliegt hierbei folgenden Annahmen. Bitte kreuzen Sie die richtige(n) Antwort(en) an:

- a) Die auftretenden Ereignisse sind voneinander unabhängig.
- b) Die Wahrscheinlichkeit von zwei oder mehr Ereignissen ist vernachlässigbar klein wenn das Intervall gross ist.
- c) Die Wahrscheinlichkeit eines Ereignisses in einem beliebig kleinen Zeitintervall $[t, t + h[$ ist proportional zu h für alle t .

(2 Punkte)

Wie hoch ist die Wahrscheinlichkeit von genau 3 Grossbränden im Kanton Zug im nächsten Jahr?

- a) 0.176
- b) 0.224
- c) 0.498

Lösung: $P(X = 3) = \frac{(3 \cdot 1)^3}{3!} e^{-3 \cdot 1} = 0.224$

1.2 (2 Punkte)

Kreuzen Sie im Folgenden die Aussage(n) an, welche für einen kontinuierlichen Zufallsprozess zutrifft (zutreffen):

- a) Ein Zufallsprozess wird als streng stationär bezeichnet, wenn ausschliesslich sein erstes und sein zweites Moment invariant über die Zeit sind.
- b) Ein Zufallsprozess wird als streng stationär bezeichnet, wenn sein erstes und sein zweites Moment und zusätzlich alle anderen Momente invariant sind über ein beliebig kleines Zeitintervall $[t, t + h[$ für alle t .
- c) Ein Zufallsprozess wird als streng ergodisch bezeichnet, wenn er streng stationär ist und zusätzlich alle seine Momente aufgrund einer Realisation des Prozesses bestimmt werden können.

1.3 (2 Punkte)

Der Mittelwert eines kontinuierlichen stochastischen Prozesses X zur Zeit t ist gegeben durch:

- a) $\mu_X(t) = \int_{-\infty}^{\infty} x f_X(x;t) dx$
- b) $\mu_X(t) = \int_{-\infty}^{\infty} x^i f_X(x;t) dx ; i = 2$
- c) Die Autokorrelationsfunktion $R_{XX}(t_1, t_2)$ für $t_1 = t_2 = t$

1.4 (2 Punkte)

Kreuzen Sie im Folgenden die Aussage(n) an, welche für eine t -Verteilung zutrifft (zutreffen):

- a) Die Varianz einer t -verteilten Zufallsvariable wird mit steigender Anzahl von Freiheitsgraden kleiner.
- b) Die t -Verteilung konvergiert für eine kleiner werdende Stichprobenzahl zu einer χ^2 -Verteilung bei gleichbleibender Anzahl an Freiheitsgraden.
- c) Die Varianz der t -Verteilung ist für einen kleinen Stichprobenumfang grösser als die Varianz der Standardnormalverteilung und nähert sich für einen großen Stichprobenumfang der Varianz der Standardnormalverteilung an.

1.5 (2 Punkte)

Ein Ingenieurbüro ist beauftragt, die Schweissnähte an einer Stahlkonstruktion hinsichtlich auftretender Risse zu überprüfen. Es soll nachgewiesen werden, dass der dabei gemessene Stichprobenmittelwert der sichtbaren Risslängen dem für die Bemessung angenommenen Mittelwert entspricht.

Für diese Überprüfung stellt das Ingenieurbüro folgende Null-Hypothese auf, welche auf einem Signifikanzniveau α von 5% getestet werden soll:

H_0 : Die Mittelwerte sind gleich.

Welche der folgenden Aussagen ist/sind richtig?

- a) Das Signifikanzniveau α entspricht allgemein der Wahrscheinlichkeit, dass die Null-Hypothese verworfen wird, obwohl sie zutrifft.
- b) Die Wahrscheinlichkeit eines Fehlers vom Typ II ist gleich 5%.
- c) Wenn das Ingenieurbüro fälschlicherweise die Null-Hypothese akzeptiert, macht es einen Fehler vom Typ I.

1.6 (2 Punkte)

Aus Erfahrungen weiss man, dass die Druckfestigkeit von Beton eines bestimmten Betonwerkes bei normaler Produktion eine Varianz von 3.54 MPa^2 besitzt. Auf einer Baustelle, die diesen Beton verwendet, wird täglich eine Stichprobe mit $n = 55$ Messungen entnommen, um die Druckfestigkeit zu beurteilen.

Geben Sie das 99% Konfidenzintervall für den wahren Mittelwert μ_x relativ zu den gemessenen Mittelwerten \bar{X} an.

- a) $\mu_x = \bar{X} \pm 0.44 \text{ MPa}$
- b) $\mu_x = \bar{X} \pm 0.66 \text{ MPa}$
- c) $\mu_x = \bar{X} \pm 0.33 \text{ MPa}$

Lösung:

$$\begin{aligned}\alpha &= 0.01 \\ k_{\alpha/2} &= 2.586 \\ n &= 55 \\ \sigma_x &= \sqrt{3.54} = 1.88 \text{ MPa} \\ \mu_x &= \bar{X} \pm 2.586 \frac{1.88}{\sqrt{55}} \\ &= \bar{X} \pm 0.66 \text{ MPa}\end{aligned}$$

- 1.7 An einer Betonbrücke soll untersucht werden, ob die Oberflächen-Chlorid-Konzentration, gemessen als Gewichtsprozent des Chlorids im Zement, mit den vorher getroffenen Annahmen in der Bemessung dieser Brücke übereinstimmt. Als Bemessungsgrundlage wurde zur Repräsentation der Chlorid-Konzentration eine normalverteilte Zufallsvariable mit einem Mittelwert μ_x von 0.42%, und einer Standardabweichung σ_x von 0.05% angenommen. Es wird nun eine Stichprobe mit Stichprobenumfang $n = 10$ entnommen, um diese Annahmen zu bestätigen.

(2 Punkte)

In welchem Intervall muss der beobachtete Mittelwert \bar{X} der Stichproben liegen, um die Annahmen auf einem Signifikanzniveau von $\alpha = 0.1$ zu bestätigen?

a) $P(0.394 \leq \bar{X} \leq 0.446) = 0.90$



b) $P(0.377 \leq \bar{X} \leq 0.455) = 0.90$



c) $P(0.388 \leq \bar{X} \leq 0.467) = 1 - \alpha$



Lösung:

Testen des Mittelwertes mit bekannter Varianz

$$P(0.42 - \Delta \leq \bar{X} \leq 0.42 + \Delta) = 1 - 0.1$$

$$\Phi\left(\frac{x_U - \mu}{\sigma}\right) - \Phi\left(\frac{x_L - \mu}{\sigma}\right) =$$

$$\Phi\left(\frac{(0.42 + \Delta) - 0.42}{\frac{0.05}{\sqrt{10}}}\right) - \Phi\left(\frac{(0.42 - \Delta) - 0.42}{\frac{0.05}{\sqrt{10}}}\right) = 0.9$$

$$\Phi\left(\frac{(0.42 + \Delta) - 0.42}{\frac{0.05}{\sqrt{10}}}\right) = 0.95 \quad (\text{nur eine Seite, daher } \alpha/2)$$

$$\Delta = 0.02632$$

(2 Punkte)

Wie gross müsste der Stichprobenumfang sein, um die Konfidenz, dass der Mittelwert im gleichen Intervall liegt, von 90% auf 99% zu erhöhen?

- a) 15
- b) 20
- c) 25

Lösung:

Konfidenzintervall $[0.394 \leq \bar{X} \leq 0.446]$

$$P\left[\mu_X - k_{\alpha/2}\sigma_X \frac{1}{\sqrt{n}} \leq \bar{X} \leq \mu_X + k_{\alpha/2}\sigma_X \frac{1}{\sqrt{n}}\right] = 1 - \alpha$$

$$\mu_X - k_{\alpha/2}\sigma_X \frac{1}{\sqrt{n}} = 0.394 \quad (\text{eine Seite gleichsetzen mit Intervallgrenze})$$

$$0.42 - 2.586 \cdot \frac{0.05}{\sqrt{n}} = 0.394$$

$$n = 24.73$$

1.8 (2 Punkte)

Welche der folgenden Aussagen ist/sind richtig?

- a) Das Wahrscheinlichkeitspapier kann verwendet werden, um zu beurteilen, ob die Datensätze zweier gleichzeitig erfasster Messgrößen linear miteinander in Zusammenhang stehen.
- b) Das Wahrscheinlichkeitspapier kann verwendet werden, um zu beurteilen, ob eine gewählte Verteilungsfunktion die beobachteten Daten ausreichend repräsentiert.
- c) Das Wahrscheinlichkeitspapier wird so konstruiert, dass die Wahrscheinlichkeitsdichtefunktion die Form einer Geraden hat.
- d) Das Wahrscheinlichkeitspapier wird so konstruiert, dass die kumulative Verteilungsfunktion die Form einer Geraden hat.

1.9 (2 Punkte)

Ein Ingenieur möchte anhand von beobachteten Daten die Verteilungsparameter μ_x und σ_x einer normalverteilten Zufallsvariable unter Verwendung der Maximum Likelihood Methode (MLM) ermitteln. Welche der folgenden Aussagen ist/sind richtig?

- a) Bei der MLM werden die Parameter einer Verteilung so gewählt, dass sie die Likelihoodfunktion minimieren.
- b) Unter Verwendung der MLM werden Punktschätzungen von Verteilungsparametern ermöglicht.
- c) Bei der MLM werden epistemische wie auch aleatorische Unsicherheiten mit der Schätzung der Verteilungsparameter berücksichtigt.

1.10 (2 Punkte)

Unter Verwendung der in der ersten Vorlesung erhobenen Daten soll ein Modell für das Gewicht der Studenten erstellt werden. Als erstes soll abgeschätzt werden, welche Verteilungsfunktion am besten diese Daten repräsentiert. Für diesen Zweck wurden die Daten in folgende Wahrscheinlichkeitspapiere (Abbildung 1.1) eingetragen.

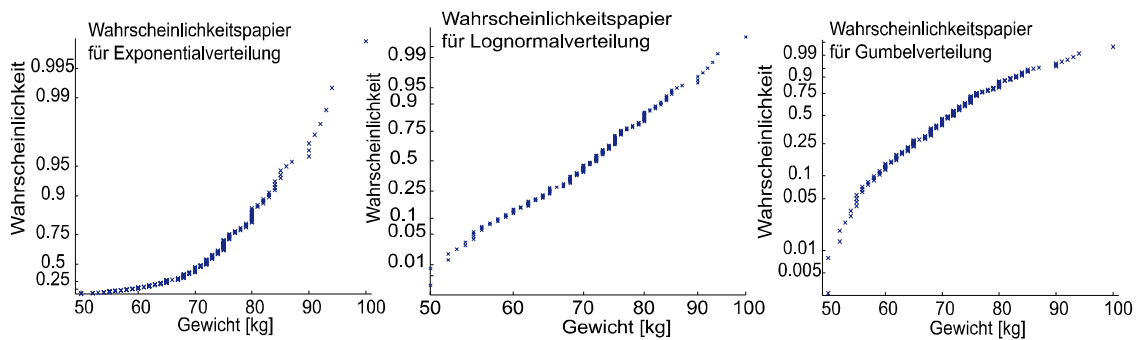


Abb. 1.1: Wahrscheinlichkeitspapier für verschiedene Verteilungen, mit den eingetragenen Daten des Körpergewichts.

Welche Verteilung repräsentiert die Daten über den gesamten Wertebereich am besten?

- a) Exponentialverteilung
- b) Lognormalverteilung
- c) Gumbelverteilung
- d) Anhand dieser Wahrscheinlichkeitspapiere lässt sich keine Aussage machen.

1.11 (2 Punkte)

Lesen Sie für die in Aufgabe 1.10 gewählte Verteilungsfunktion den Median aus der Abbildung 1.1 ab. Der Median ist:

- a) 82 kg
- b) 72 kg
- c) 65 kg
- d) 0.9 kg

Name:

1.12 (2 Punkte)

Das Gewicht der Studierenden soll durch eine normalverteilte Zufallsvariable X repräsentiert werden. Schätzen Sie den Mittelwert μ_X und die nichterwartungstreue Standardabweichung σ_X mittels der Methode der Momente. Verwenden Sie dazu die Stichprobenwerte aus Tabelle 1.1.

Tab. 1.1: Körpergewicht, zufällige Stichprobe von 10 Studierenden.

i	x_i : Körpergewicht [kg]
1	64
2	65
3	66
4	68
5	69
6	70
7	74
8	76
9	78
10	81

- a) $\hat{\mu}_X = 71.1; \hat{\sigma}_X = 30.7$
- b) $\hat{\mu}_X = 70.9; \hat{\sigma}_X = 5.6$
- c) $\hat{\mu}_X = 71.1; \hat{\sigma}_X = 5.84$
- d) $\hat{\mu}_X = 71.1; \hat{\sigma}_X = 5.5$

Lösung:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = 71.1$$

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \hat{\mu}^2} = 5.5$$

1.13 (2 Punkte)

Ein χ^2 -Anpassungstest soll durchgeführt werden um festzustellen, wie gut das Modell (bestehend aus Verteilungsfunktion und Parameter) aus Aufgabe 1.12 die Daten repräsentiert. Dafür teilen wir unsere Daten aus Tabelle 1.1 in vier Klassen ein. Wie viele Freiheitsgrade hat die χ^2 -Verteilung?

- a) 0
- b) 1
- c) 2
- d) 3

(Lösung: Freiheitsgrade = Klassen - 1 Anzahl - Anzahl mit diesen Beobachtung bereits abgeschätzten Parameter = 4 - 1 - 2 = 1)

(1 Punkt)

Genügt die Anzahl der Beobachtungen, um ein aussagekräftiges Resultat für den χ^2 -Test zu erhalten?

- a) Ja, die Anzahl der Beobachtungen reicht aus.
- b) Nein, die Anzahl der Beobachtungen ist zu klein.

(Lösung: Nein, pro Klasse sollten mindestens 5 Daten sein → da wir nur 10 Werte in der Tabelle haben reicht das nicht... Skript Seite E-30)

Teil 2: Rechenaufgaben (maximal 27 Punkte)

2.1 Der Kreisingenieur einer Wintersportregion will die jährliche maximale Schneehöhe mit Hilfe einer gumbelverteilten Zufallsvariable X mit einem Mittelwert von $\mu_x = 4.5$ m und einer Varianz von $\sigma_x^2 = 1.4$ m² modellieren.

a) (3 Punkte)

Ermitteln Sie die kumulative Verteilungsfunktion für die 25-jährige maximale Schneehöhe, unter Angabe ihrer Verteilungsparameter. Es wird angenommen, dass die jährlichen Maxima voneinander unabhängig sind.

b) (3 Punkte)

Wie gross ist die Schneehöhe, die einer erwarteten Wiederkehrperiode von 75 Jahren entspricht?

c) (3 Punkte)

Berechnen Sie die Wahrscheinlichkeit, dass die jährliche maximale Schneehöhe 5 m übersteigt.

Hinweis: Die Gumbelverteilung (Gumbel max) hat die nachfolgende Form:

$$-\infty < x < \infty$$

$$F_X(x) = \exp(-\exp(-\alpha(x-u)))$$

$$\mu_x = u + \frac{0.577216}{\alpha}$$

$$\sigma_x = \frac{\pi}{\alpha\sqrt{6}}$$

μ_x – Mittelwert

σ_x – Standardabweichung

u – Parameter der Verteilung

α – Parameter der Verteilung

Lösung:

a)

$$F_Y(y) = P[Y \leq y] = [F_X(x)]^{25} = [\exp(-\exp(-\alpha(x-u)))]^{25} = \exp(-25 \exp(-\alpha(x-u)))$$

Parameter u und α :

$$\mu_x = u + \frac{0.577216}{\alpha}$$

$$\alpha = \frac{\pi}{\sigma_x \sqrt{6}} = \frac{\pi}{\sqrt{1.4} \sqrt{6}} = 1.084$$

$$\sigma_x = \frac{\pi}{\alpha \sqrt{6}}$$

$$u = \mu_x - \frac{0.57722}{\alpha} = 4.5 - \frac{0.57722}{1.084} = 3.97$$

b)

Jährliche Auftretenswahrscheinlichkeit: $p = \frac{1}{T} = \frac{1}{75}$

$$1 - P[\text{jährliches Max} > x] = F_x(x)$$

$$1 - \frac{1}{75} = F_x(x)$$

$$1 - \frac{1}{75} = F_x(x) = \exp(-\exp(-\alpha(x-u))) = 0.987$$

$$\Leftrightarrow \ln(-\ln(0.987)) = -\alpha(x-u) \Leftrightarrow \frac{\ln(-\ln(0.987))}{-\alpha} + u = x$$

$$\Leftrightarrow \frac{\ln(-\ln(0.987))}{-1.084} + 3.97 = x \Leftrightarrow 4.00 + 3.97 = x$$

$$\Leftrightarrow x = \underline{\underline{7.97m}}$$

c)

$$F_x(x) = \exp(-\exp(-\alpha(x-u)))$$

$$P[\text{jährliches Max} > 5m] = 1 - F_x(x=5) = 1 - \exp(-\exp(-\alpha(5-u)))$$

$$1 - F_x(x=5) = 1 - \exp(-\exp(-1.084(5-3.97))) \quad \alpha = 1.084$$

$$= 1 - \exp(-\exp(-1.117)) = 1 - 0.721 = 0.279 \quad u = 3.97$$

Die Wahrscheinlichkeit, dass die jährliche maximale Schneehöhe 5m überschreitet, beträgt 0.279.

2.2 (6 Punkte)

Gegeben ist in der folgenden Tabelle 2.1 eine Stichprobe mit Stichprobenumfang $n = 5$ der Körpergewichte $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)^T$ der Studierenden im 2. Semester. Das Körpergewicht wird durch die Zufallsvariable X repräsentiert. Es wird eine passende Wahrscheinlichkeitsverteilung für das Körpergewicht gesucht. Dazu soll mit einem Kolmogorov-Smirnov Test geprüft werden, ob das Körpergewicht durch eine Normalverteilung mit den Parametern $\mu_x = 70$ kg und $\sigma_x = 9$ kg repräsentiert werden kann. Für die Lösung der folgenden Aufgaben stehen Ihnen entsprechende Tabellen im Anhang der Prüfung zur Verfügung.

Tabelle 2.1: Stichprobenstatistik des Körpergewichts der Studierenden

i	Körpergewicht x_i^o , sortierte Beobachtungen	$F_o(x_i^o) = \frac{i}{n}$	$F_p(x_i^o)$	$ F_o(x_i^o) - F_p(x_i^o) $
1	64	0.2	$\frac{64 - 70}{9} = -0.67$; $\Phi(0.67) = 0.7486$; $\Phi(-0.67) = 1 - 0.7486$ $= 0.2514$	$ 0.2 - 0.25 = 0.05$
2	66	0.4	$\Phi(-0.44) = 0.33$	0.07
3	69	0.6	$\Phi(-0.11) = 0.46$	0.14
4	74	0.8	$\Phi(0.44) = 0.67$	0.13
5	78	1.0	$\Phi(0.89) = 0.81$	0.19

Füllen Sie bitte die Tabelle vollständig aus.

Führen Sie einen Kolmogorov-Smirnov-Test für die Güte der Anpassung auf einem Signifikanzniveau von 10% durch, um zu testen, ob aufgrund dieser Daten die Körpergröße durch eine Normalverteilung mit den oben gegebenen Parametern repräsentiert werden kann.

Für $n = 5, \alpha = 10\%$ ist der kritische Kolmogorov-Smirnov-Wert $= 0.509$.

Da $0.509 > 0.190$, ist die Normalverteilung mit den gegebenen Parametern für diese Daten akzeptierbar.

2.3 (6 Punkte)

An der EPF Lausanne wurden von 100 Studenten ebenfalls die Körpergewichte erfasst. Es wird davon ausgegangen, dass der wahre Mittelwert der Gewichte der Studenten der ETH mit $\mu_{ETHZ} = 71$ kg bekannt ist.

Die gemessenen Werte der Gewichte der Studenten der EPFL haben einen Stichprobenmittelwert von $\bar{x}_{EPFL} = 75$ kg und eine Stichprobenstandardabweichung von $s_{EPFL} = 12.3$ kg.

Es soll auf einem Signifikanzniveau von $\alpha = 0.1$ überprüft werden, ob die beiden Gruppen aus der gleichen Grundgesamtheit stammen, oder ob eine der beiden Studentengruppen signifikant schwerer oder leichter ist.

Benützen Sie zur Lösung der Aufgabe die entsprechenden Tabellen im Anhang.

Lösung:

- 1) Erkennen, dass ein t-Test verwendet werden muss
- 2) Festlegen der Nullhypothese und Auswahl des Signifikanzniveaus

$$\begin{aligned} H_0 : \mu_{ETHZ} &= \mu_{EPFL} & \alpha &= 0.1 \\ H_1 : \mu_{ETHZ} &\neq \mu_{EPFL} \end{aligned}$$

- 3) Festlegen der Entscheidungsregel

Der Stichprobenmittelwert muss in einem zu bestimmenden Intervall rund um den wahren Mittelwert der Gewichte der Studenten der ETH μ_{ETHZ} liegen.

$$P[-t_{\alpha/2} \leq T = \frac{\bar{X}_{EPFL} - \mu_{ETHZ}}{S_{EPFL} / \sqrt{n}} \leq t_{\alpha/2}] = 1 - \alpha \quad n = 100$$

- 4) Berechnung der Akzeptanzkriterien:

$$\mu_{ETHZ} - t_{\alpha/2} \frac{s_{EPFL}}{\sqrt{n}} \leq \bar{x}_{EPFL} \leq \mu_L + t_{\alpha/2} \frac{s_{EPFL}}{\sqrt{n}}$$

$$\text{Untere Intervallgrenze: } \mu_L - t_{\alpha/2} \frac{s_M}{\sqrt{n}} = 71 - 1.66 \cdot \frac{12.3}{\sqrt{100}} = 68.9582$$

$$\text{Obere Intervallgrenze: } \mu_L + t_{\alpha/2} \frac{s_M}{\sqrt{n}} = 71 + 1.66 \cdot \frac{12.3}{\sqrt{100}} = 73.0418$$

- 5) Durchführung des Tests

$$\bar{x}_{EPFL} = 75 \text{ kg}$$

Das Intervall der t-Statistik wurde berechnet als [68.96kg; 73.04kg].

Der Stichprobenmittelwert liegt nicht im Intervall.

- 6) Rückschluss auf gewähltem Signifikanzniveau

Die Nullhypothese, dass es sich um die gleiche Grundgesamtheit handelt, wird auf einem Signifikanzniveau von 10% verworfen.

2.4 In der Basler Region ist durchschnittlich ein Erdbeben pro Jahr zu beobachten, das eine Intensität (gemessen in Richter-Magnitude) von mehr als 3.0 aufweist.

a) (3 Punkte)

Wie gross ist die Wahrscheinlichkeit, dass in dieser Region in irgendeinem beliebigen Jahr ein Erdbeben mit einer Intensität grösser als 3.0 auftritt?

$$P[T \leq 1 \text{ Jahr}] = 1 - e^{-\nu \cdot t} = 1 - e^{-\frac{1}{1}} = 1 - e^{-1} = 63.2 \%$$

Es wird angenommen, dass die Richter-Magnitude von Erdbebenereignissen mit einer Exponentialverteilung beschrieben werden kann. In der Bergregion des Kantons Wallis wurde der Parameter dieser Verteilung auf $\lambda = 2.35$ geschätzt.

b) (3 Punkte)

Wie gross ist die Wahrscheinlichkeit, dass irgendein Erdbeben in dieser Region einen Wert grösser als 6.3 erreichen wird?

$$f_M(m) = \lambda e^{-\lambda m} \quad (\text{Exponentialverteilung})$$

mit $\lambda = 2.35$

$$f_M(m) = 2.35 e^{-2.35m}$$
$$F_M(m) = 1 - e^{-2.35m}$$

$$P[M \geq 6.3] = 1 - F_M(6.3) = e^{-(2.35)6.3} = e^{-14.8} = 3.72 \cdot 10^{-7}$$