Schriftliche Prüfung
(2 Stunden)

Bemerkungen:

- Erlaubte Hilfsmittel: 10 hand- oder maschinengeschriebene A4 Seiten (=5 Blätter). Taschenrechner ohne Kommunikationsmöglichkeit.
- Mobiltelefone sind auszuschalten!
- Lesen Sie zuerst alle Aufgaben durch! Verweilen Sie nicht zu lange bei einem Aufgabenteil, der Ihnen grosse Schwierigkeiten bereitet!
- Die nötigen Tabellen befinden sich auf den hintersten Seiten dieser Prüfung.
- Der Lösungsweg muss klar ersichtlich sein (Ausnahme: Multiple-Choice Aufgaben). Wenn Ihnen für einen Aufgabenteil ein Zwischenresultat fehlt, treffen Sie eine sinnvolle Annahme und markieren Sie diese deutlich.
- Aufgaben 1 bis 4 sind Textaufgaben. Schreiben Sie Ihre Antworten zu diesen Aufgaben auf die dafür vorgesehenen Antwortblätter und nicht auf das Aufgabenblatt.
- Aufgabe 5 ist eine Multiple-Choice Aufgabe. Jede Teilaufgabe besteht aus mehreren Aussagen. Pro Teilaufgabe können keine, eine oder mehrere Aussagen richtig sein. Die Antworten müssen auf dem separaten Multiple-Choice Antwortblatt (blau) am Ende des Antwortdossiers angekreuzt werden und dürfen *nicht* auf der Prüfung angekreuzt werden. Für jede Aussage gibt es 1/2 Punkt, wenn sie korrekt markiert wird. 1/2 Punkt wird abgezogen, wenn eine Aussage falsch markiert wird. Wenn eine Aussage nicht markiert wird, gibt es keinen Punkt, es wird aber auch kein Punkt abgezogen. Keine Teilaufgabe kann mit negativer Punktzahl abgeschlossen werden. Minimal erhält man also für eine ganze Teilaufgabe 0 Punkte.

Viel Erfolg!

1. (8 Punkte)
Anna isst gerne Joghurts und hat jederzeit viele Joghurts im Kühlschrank stehen. Manchmal verliert sie den Überblick und ein paar Joghurts überschreiten das Haltbarkeitsdatum um wenige Tage. Aus Erfahrung passiert ihr dies in 10% der Fällen ($\mathbb{P}(H^c) = 0.1$). Da Anna aber nicht gerne Lebensmittel wegwirft, möchte sie die Situation statistisch anschauen. Sie fragt die Joghurtfirma um Auskunft, welche ihr verspricht, dass beim Einhalten des Haltbarkeitsdatums sich in 99.5% der Fällen auch keine Anzeichen von Schimmel bilden ($\mathbb{P}(S^c \mid H) = 0.995$). Anna weiss aus Erfahrung auch, dass Joghurts, welche bei ihr im Kühlschrank das Haltbarkeitsdatum überschreiten, oft noch keinen Schimmel gebildet haben. Sie rechnet mit 70% ($\mathbb{P}(S^c \mid H^c) = 0.7$).

Wir definieren folgende Ereignisse:

$$\begin{aligned} H &:= \text{Haltbarkeitsdatum nicht überschritten} \\ H^c &:= \text{Haltbarkeitsdatum überschritten} \\ S &:= \text{Schimmelbildung (Joghurt ungeniessbar)} \\ S^c &:= \text{keine Anzeichen von Schimmel (Joghurt geniessbar)} \end{aligned}$$

a) (1 Punkt) Wie hoch ist die Wahrscheinlichkeit, dass bei einem Joghurt mit eingehaltenem Haltbarkeitsdatum sich trotzdem Schimmel bildet?

b) (1 Punkt) Wie hoch ist die Wahrscheinlichkeit, dass Anna einen Joghurt aus ihrem Kühlschrank herauszieht, welcher die Haltbarkeit nicht überschreitet und keinen Schimmel hat?

c) (2 Punkte) Thomas ist bei Anna zu Besuch und möchte einen Joghurt essen. Mit welcher Wahrscheinlichkeit nimmt Thomas einen schimmeligen Joghurt aus Annas Kühlschrank? (Ersatzresultat: 0.04)

d) (3 Punkte) Wie hoch ist die Wahrscheinlichkeit, dass ein schimmelfreier Joghurt in Annas Kühlschrank das Haltbarkeitsdatum einhält ($\mathbb{P}(H \mid S^c)$)? Sind H und S^c unabhängig?

e) (1 Punkt) Was ist die Wahrscheinlichkeit $\mathbb{P}(H \mid H)$?

2. (8 Punkte)
Der Forscher Dominik versucht in seinen Ferien (zum Leidwesen seiner Frau) die Verteilung von Sandkorngrössen in der Nähe des Strandhauses zu untersuchen. Mit Hilfe eines Mikroskops hat er folgende kumulative Verteilungsfunktion bestimmt

$$F(x) := \frac{x^2}{x^2+1} \quad \left(= \frac{1}{1+\frac{1}{x^2}} \right), x > 0$$

welche den Durchmesser X (in mm) von Sandkörnern sinnvoll beschreibt.

a) (1 Punkt) Bestimmen Sie die Dichte zur Verteilungsfunktion $F(x)$.

b) (1.5 Punkte) Zeigen Sie, dass für das $\alpha-$Quantil q_α gilt

$$q_\alpha = \sqrt{\frac{\alpha}{1-\alpha}}.$$

c) (1 Punkt) Wie hoch ist die Wahrscheinlichkeit, dass Dominik ein Sandkorn in die Hand nimmt, welches grösser als $1mm$ ist?

d) (1 Punkt) Aus seinen Messungen hat Dominik beobachtet, dass etwas mehr als die Hälfte der Sandkörner einen Durchmesser zwischen $0.5mm$ und $2mm$ haben. Was ist die Wahrscheinlichkeit unter F, dass ein Sandkorn einen Durchmesser zwischen $0.5mm$ und $2mm$ hat?

e) (1.5 Punkte) Angenommen wir beobachten die Durchmesser von zwei Sandkörnern (Unabhängigkeit ist gegeben). Wie hoch ist die Wahrscheinlichkeit, dass beide Durchmesser kleiner als $1mm$ sind?

f) (2 Punkte) Wie viele zufällige, unabhängige Sandkörner muss Dominik mindestens unter dem Mikroskop anschauen, um mit einer Wahrscheinlichkeit von mehr als 90% mindestens ein Sandkorn mit einem Durchmesser grösser als $2mm$ zu messen?

3. (**7 Punkte**)
Jan ist Versicherungsexperte. Um die Kosten X (in CHF) von Versicherungsfällen zu modellieren, verwendet er folgende Dichtefunktion:

$$f(x) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\theta\pi}} \exp\left(-\frac{x^2}{2\theta}\right) & \text{für } x > 0 \\ 0 & \text{für } x \leq 0, \end{cases}$$

wobei $\theta > 0$ ein unbekannter Modellparameter ist.

a) (1 Punkt) Bestimmen Sie die Likelihoodfunktion basierend auf n unabhängigen, identisch verteilten Beobachtungen $x_1, \ldots, x_n$ einer Zufallsvariable mit obiger Dichte. Vereinfachen Sie die Terme jeweils so weit wie möglich.

b) (2 Punkte) Bestimmen Sie den Maximum-Likelihood-Schätzer $\hat{\theta}_{\text{MLE}}$ für θ basierend auf n unabhängigen, identisch verteilten Beobachtungen $x_1, \ldots, x_n$. Schreiben Sie die allgemeine Form auf. (Ersatzresultat: $\hat{\theta}_{\text{MLE}} = \frac{1}{n\pi} \sum_{i=1}^{n} x_i^2$)

c) (2 Punkte) Bestimmen Sie die Varianz von $\hat{\theta}_{\text{MLE}}$ für den Fall $\theta = 1$. Sie können verwenden, dass $\text{Var}(X^2) = 2\theta^2$.

d) (2 Punkte) Berechnen Sie $\mathbb{E}(X)$ (Ersatzresultat: $\mathbb{E}(X) = \frac{\sqrt{2\theta}}{\sqrt{3}}$) und bestimmen Sie damit den Momentenschätzer $\hat{\theta}_{\text{Moment}}$ für θ basierend auf n unabhängigen, identisch verteilten Beobachtungen $x_1, \ldots, x_n$.

4. (7 Punkte)

Ein Lebensmittelhersteller behauptet, dass ein neuentwickeltes Proteinbrot seiner Marke deutlich mehr Eiweiss enthält als das der Konkurrenz. Es soll 40g Eiweiss pro 100g Brot enthalten. Der Alpinist Christopher ist skeptisch und misst den Eiweissgehalt in $n = 30$ Broten nach. Die Brote haben jeweils eine Masse von 100g. Wir nehmen an, dass die Messungen unabhängig voneinander und identisch gemäss einer Normalverteilung verteilt sind. Mit X_i bezeichnen wir die i-te Messung des Eiweissgehaltes. Wir haben also $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Folgende Schätzer wurden ermittelt:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = 39.14$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu})^2 = 34.44$$

Wir wollen nun mit einem t-Test untersuchen, ob die Messergebnisse auf einen signifikant kleineren Wert für den Eiweissgehalt hinweisen als vom Hersteller angegeben. Verwenden Sie das Signifikanzniveau $\alpha = 0.05$.

a) (2 Punkte) Schreiben Sie die Nullhypothese, die Alternativhypothese und die Teststatistik dazu auf. Um welche Art von Test handelt es sich? (Einstichprobentest, Zweistichprobentest, einseitig, zweiseitig, ...)

b) (1 Punkt) Führen Sie den entsprechenden t-Test durch: Was ist der Wert der Teststatistik? Was ist das Resultat des Tests?

Nehmen Sie nun an, dass die Standardabweichung bekannt ist und verwenden Sie $\sigma = 5$. Wir wollen weiterhin untersuchen, ob die Messergebnisse auf einen signifikant kleineren Wert für den Eiweissgehalt hinweisen als vom Hersteller angegeben.

c) (2 Punkte) Betrachten Sie nun den entsprechenden Z-Test mit $\alpha = 0.05$. Welchen Wert müssten wir für das arithmetische Mittel $\hat{\mu}$ beobachten, um die Nullhypothese gerade noch zu verwerfen? Runden Sie das Ergebnis auf zwei Nachkommastellen.

d) (2 Punkte) Berechnen Sie die Macht des Z-Tests unter der Alternativhypothese $\mu_A = 39$ mit $\sigma = 5$. Runden Sie das Ergebnis auf zwei Nachkommastellen.

5. (**10 Punkte**) Entscheiden Sie für jede Aussage, ob sie richtig oder falsch ist. Markieren Sie Ihre Antwort auf dem separaten Multiple Choice Antwortblatt.

a) (2 Punkte) Seien $X \sim \mathcal{N}(0, \sigma_X^2)$ und $Y \sim \mathcal{N}(0, \sigma_Y^2)$ mit $\mathrm{Cov}(X, Y) = 0$. Dann gilt:

1. X und Y sind unkorreliert.
2. $\mathbb{E}[(X+Y)^2] = \sigma_X^2 + \sigma_Y^2$.
3. Falls $\sigma_Y > \sigma_X$, dann ist das 99%-Quantil von Y grösser als das 99%-Quantil von X.
4. $\mathbb{P}(X > x) = \Phi\left(-\frac{x}{\sigma_X}\right)$, wobei Φ die kumulative Verteilungsfunktion der Standardnormalverteilung ist.

b) (2 Punkte) Seien $X \in \{0, 1\}$ und $Y \in \{0, 1\}$ zwei binäre Zufallsvariablen. In untenstehender (unvollständiger) Tabelle stehen die Wahrscheinlichkeiten für die vier möglichen Fälle, die eintreten können.

	$X = 0$	$X = 1$
$Y = 0$	1/3	1/6
$Y = 1$	1/4	??

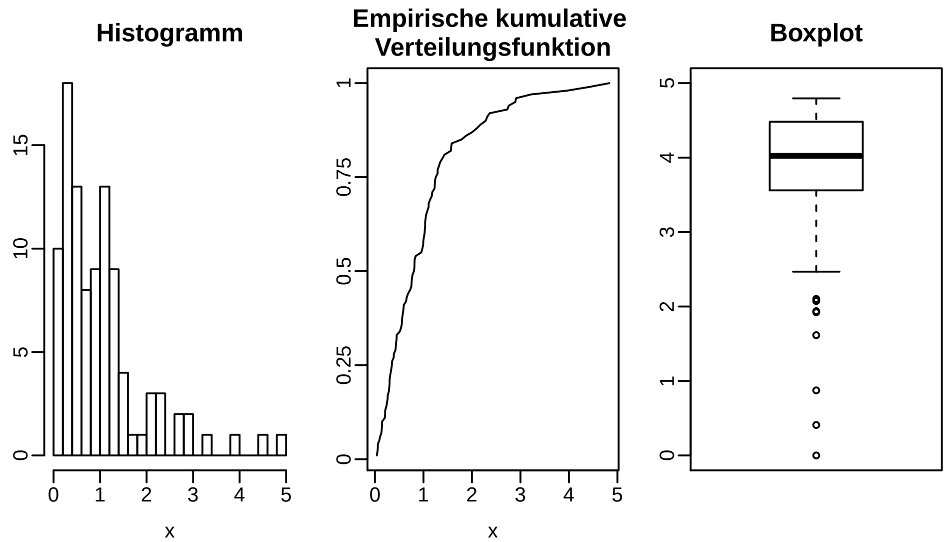
Beurteilen Sie die folgenden Aussagen:

1. $\mathbb{P}(X = 1, Y = 1) = \frac{1}{4}$
2. $\mathbb{P}(Y = 0) = \mathbb{P}(X = 0)$
3. $\mathbb{E}[XY] = \mathbb{P}(X = 1, Y = 1)$
4. Binäre Zufallsvariablen sind immer Bernoulli-verteilt.

c) (2 Punkte) Entscheiden Sie für jede der folgenden Aussagen, ob sie richtig oder falsch ist.

1. Der zentrale Grenzwertsatz besagt, dass die Summe beliebiger i.i.d.-verteilter Zufallsvariablen $X_1, ..., X_n$ für grosses n approximativ standardnormalverteilt ist.
2. Seien $X_1, ..., X_n$ i.i.d. $\mathrm{Pois}(\lambda)$. Dann gilt:
$$S_n = \sum_{i=1}^{n} X_i \sim \mathrm{Pois}(n\lambda) \qquad \text{und} \qquad \overline{X}_n = \frac{1}{n} S_n \sim \mathrm{Pois}(\lambda).$$
3. Seien $X_1, ..., X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$ und sei $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Dann gilt für jedes beliebige n, dass
$$\mathbb{E}[\,\overline{X}_n\,] = \mu.$$
4. Für die Summe S_n von n i.i.d. $\mathrm{Exp}(\lambda)$-verteilten Zufallsvariablen gilt für n genügend gross, dass
$$\frac{S_n - \frac{n}{\lambda}}{\frac{\sqrt{n}}{\lambda}} \approx \mathcal{N}(0, 1).$$

d) (2 Punkte) Beurteilen Sie folgende Aussagen zu deskriptiver Statistik. Die ersten drei Teilaufgaben beziehen sich auf die folgenden Darstellungen von jeweils 100 Beobachtungen.



1. Der Boxplot zeigt dieselben 100 Beobachtungen wie die empirische kumulative Verteilungsfunktion.
2. Die Quartilsdifferenz der Beobachtungen in der empirischen kumulativen Verteilungsfunktion ist etwa 1.
3. Das Histogramm passt zu der empirischen kumulativen Verteilungsfunktion.
4. Wenn man eine ungerade Anzahl Beobachtungen hat, entspricht der Median immer exakt einer der Beobachtungen.

e) (2 Punkte) Beurteilen Sie folgende Aussagen zu statistischen Tests.

1. Ein $(1-\alpha)\times 100\%$ Vertrauensintervall I für den Modellparameter θ ist definiert als

   $$I = \{\theta_0 \;:\; \text{Nullhypothese } H_0 \;:\; \theta = \theta_0 \text{ wird für Signifikanzniveau } \alpha \text{ verworfen } \}.$$

2. Bei der Berechnung eines p-Wertes geht man immer davon aus, dass die Nullhypothese stimmt.
3. Die Wahrscheinlichkeit eines Fehlers 2. Art kann nur unter Annahme eines konkreten Parameters $\theta_A \in H_A$ berechnet werden.
4. Durch Sammeln von zusätzlichen Beobachtungen kann man die Signifikanz eines Testergebnisses nahezu beliebig erhöhen, wenn die Nullhypothese nicht stimmt. Dies bedeutet aber nicht immer, dass das Ergebnis auch fachlich relevant ist.

Tabelle der Standardnormalverteilung

$$\Phi(z) = \mathrm{P}\left[Z \leq z\right], \; Z \sim \mathcal{N}(0,1)$$

Bsp.: $\mathrm{P}\left[Z \leq 1.96\right] = 0.975$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Quantile der t-Verteilung

Bsp.: $t_{9,\,0.975} = 2.262$

df	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
31	0.255	0.530	0.853	1.309	1.696	2.040	2.452	2.744
32	0.255	0.530	0.853	1.309	1.694	2.037	2.449	2.738
33	0.255	0.530	0.853	1.308	1.693	2.035	2.445	2.733
34	0.255	0.529	0.852	1.307	1.691	2.032	2.441	2.728
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576