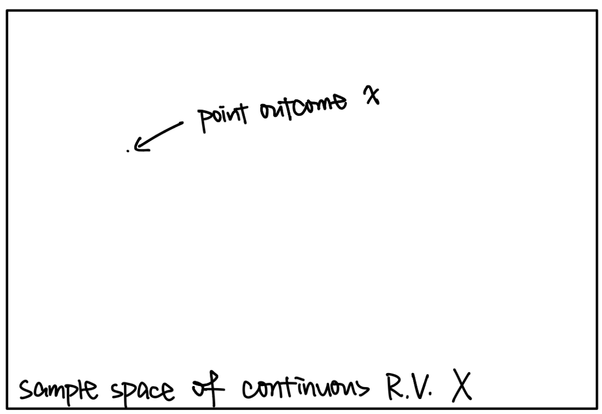
Dis 6: Continuous Probability Distributions

Related textbook chapter: 8
Ch 8 handout and solution offered by Dr. Pac can be accessed here: Handout Solution
This handout incorporates reviews with all exercises from the handout given by Dr. Pac.

1 Density Functions

- When we examined discrete probability distributions, we said that a discrete probability distribution should describe point probability $P(x)$ at all possible x outcome values.
- Ideally, we would like to provide a similar definition for continuous probability distributions.
- But there's a problem: when random variable X is continuous, **point probability equals to 0 at every single point** ($P(X = x) = 0$ for all x).
  - Reason 1: A continuous random variable has uncountable amount of values, so if each outcome value has probability $\varepsilon > 0$, then the sum of all probabilities would equal to ∞ instead of 1.
  - Reason 2: Say that the rectangle below represents the sample space for a continuous random variable. The probability of hitting a point within the rectangle is the area of the point divided by the area of the rectangle. However, a point has area $= 0$, so the point probability $= 0$.



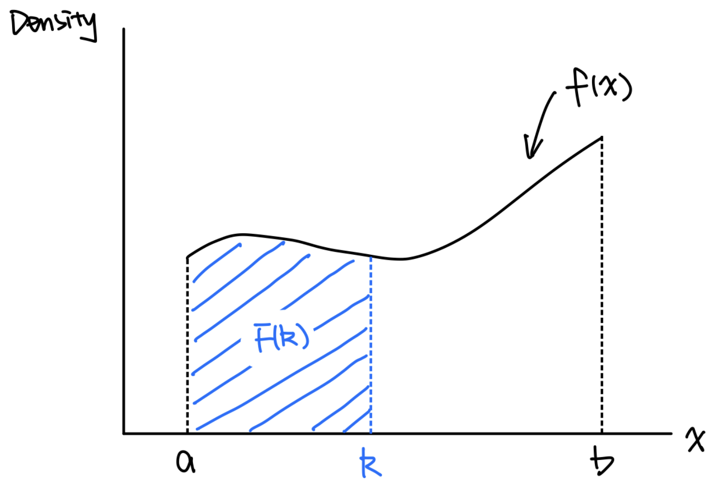
- So our definition for a continuous probability distribution needs to be modified (slightly).
- Solution: describe **density** of x instead of probability.

Definition 1 (Probability density function (PDF)). A function $f(x)$ is called a probability density function (PDF) over $a \leq x \leq b$ if it satisfies the following two criteria:

1. $f(x) \geq 0$ for all x between a and b, and
2. Total area under the curve of $f(x)$ between a and b is 1.

Definition 2 (Cumulative density function (CDF)). A cumulative density function (CDF) describes probability up to a point x. That is, CDF $F(x) = P(X \leq x)$ for random variable X.

Exercise.



1. Label $f(x)$ and $F(k)$ on the graph above.
   See the labelled graph above.
2. In order for $f(x)$ to be a PDF, what additional requirement is needed?
   Two requirements are needed for $f(x)$ to be a PDF: $f(x) \geq 0$ for all x between a and b, and the total area between a and b equals to 1.
   The first requirement is already satisfied given that $f(x)$ lies above the horizontal axis. Thus, the additional requirement needed is that $F(b) = P(X \leq b) = 1$.

- With the help of density functions, we can finally define a continuous probability distribution:

Definition 3 (Continuous probability distribution)**.** A continuous probability distribution describes a valid PDF $f(x)$ at all possible x values for a continuous random variable X.

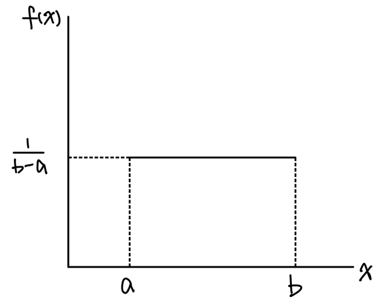
How does this compare with the discrete case?

	Discrete Prob Dist	**Continuous Prob Dist**
Describes . . . at all valid x	$P(x)$	$f(x)$
Range of measure for all valid x	$0 \leq P(x) \leq 1$	$f(x) \geq 0$
How to make sure all valid x are covered	$\sum_x P(x) = 1$	$F(b) = \int_a^b f(x)dx = 1$

2 Examples of Continuous Probability Distribution

2.1 Uniform Distribution

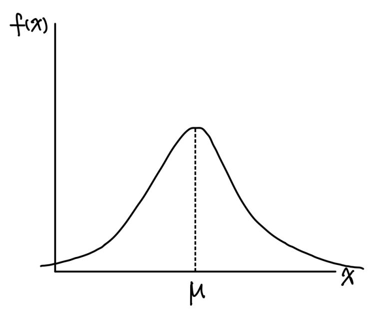
- If X is uniformly distributed between point a and b, then $X \sim \text{Uniform}(a, b)$
- PDF: $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$



- $E(X) = \frac{a+b}{2}$
- $V(X) = \frac{(b-a)^2}{12}$

2.2 Normal Distribution

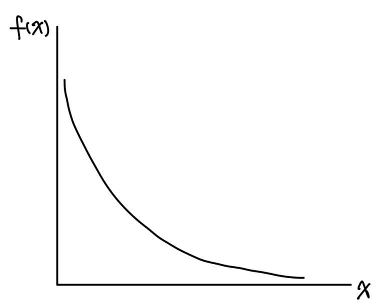
- If X is normally distributed with expected value μ and variance σ^2, then $X \sim N(\mu, \sigma^2)$
- PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ for $-\infty < x < \infty$



- Usually, for random variable X that follows a normal distribution, it is helpful to standardize the variable so that we are examining a transformed variable with **standard normal distribution** instead.
  - A random variable Z that follows standard normal distribution is denoted as $Z \sim N(0, 1)$
  - How to transform X to be standard normal? $\Rightarrow$ Since $X \sim N(\mu, \sigma^2)$, $\frac{X-\mu}{\sigma} = Z \sim N(0, 1)$

2.3 Exponential Distribution

- An exponential distribution describes the waiting time until the next "success" event.
- If X is exponentially distributed with success arrival rate λ, then $X \sim \text{exponential}(\lambda)$
- PDF: $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$



- CDF: $F(x) = P(X \leq x) = 1 - e^{-\lambda x}$ for $x \geq 0$
- $E(X) = \frac{1}{\lambda}$
- $\sigma = \sqrt{V(X)} = \frac{1}{\lambda}$

3 Exercises

1. The weekly output of a steel mill is a uniformly distributed random variable that lies between 110 and 175 metric tons

   (a) What is the probability the steel mill will produce more than 150 metric tons next week?

   Notice that this is a uniform distribution with $a = 110$ and $b = 175$, so the PDF is

   $$f(x) = \frac{1}{b-a} = \frac{1}{175-110} = \frac{1}{65}$$

   It would be perfectly valid to answer this question by integrating the PDF from 150 to the top of the sample space (175), but since this is a uniform distribution, the area under the PDF is a rectangle, so it's easier to calculate the probability by multiplying width times height. In this case:

   $$P(150 < \text{output} < 175) = (175 - 150) \times \frac{1}{65} = 0.3846$$

   (b) What is the probability the mill will produce between 120 and 160 metric tons next week?

   Since between 120 and 160 metric tons is still within the range $[a, b] = [110, 175]$ that the uniform distribution is defined on, we don't have to worry about any weird out-of-bound problem. Hence,

   $$P(120 < \text{output} < 160) = (160 - 120) \times \frac{1}{65} = 0.6154$$

   (c) What is the expected value and variance of the mill's weekly output?

   Using the formulas for the expected value and variance of a uniformly distributed random variable, we obtain

   $$E(X) = \frac{a+b}{2} = \frac{110+175}{2} = 142.5$$
   $$V(X) = \frac{(b-a)^2}{12} = \frac{(175-110)^2}{12} = 352.083$$

2. An analysis of the amount of interest paid monthly by Visa cardholders is normally distributed with a mean of \$27 and a standard deviation of \$6. (Note: A probability table for a standard normal can be found on the last page of the handout.)

   (a) What proportion of Visa cardholders pay less than \$30 in interest?

   We want $P(X \leq 30)$, where $X \sim N(27, 6^2)$. In order to make headway, we must first convert the question into one involving a standard normal. This is done by subtracting the mean and dividing by the standard deviation:

$$P(X \leq 30) = P\left(\frac{X-\mu}{\sigma} \leq \frac{30-\mu}{\sigma}\right) = P\left(\frac{X-27}{6} \leq \frac{30-27}{6}\right) = P(Z \leq 0.50)$$

Now we have a probability that we can find using our standard normal table. Looking up the probability when $z = 0.50$, we find

$$P(X \leq 30) = P(Z \leq 0.50) = 0.6915$$

Z	0.00	0.01	0.02	0.03	0.04	0.
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7

(b) What proportion pay more than \$42 in interest?
The strategy is similar to before. The only difference is that we must use the complement rule this time to flip the inequality:

$$\begin{aligned} P(X > 42) &= P\left(\frac{X-27}{6} > \frac{42-27}{6}\right) = P(Z > 2.5) \\ &= 1 - P(Z \leq 2.5) = 1 - 0.9938 = 0.0062 \end{aligned}$$

2.2	0.9861	0.9864	0.9868	0.9871	0.987
2.3	0.9893	0.9896	0.9898	0.9901	0.990
2.4	0.9918	0.9920	0.9922	0.9925	0.992
2.5	0.9938	0.9940	0.9941	0.9943	0.994
2.6	0.9953	0.9955	0.9956	0.9957	0.995

(c) What proportion pay less than \$15 in interest?
Again, we begin by standardizing the normal distribution:

$$P(X \leq 15) = P\left(\frac{X-27}{6} \leq \frac{15-27}{6}\right) = P(Z \leq -2)$$

Now the extra twist is that you are only provided with the probability table for $z \geq 0$. To make headway, we must make use of the fact that the standard normal distribution is symmetric:

$$P(Z \leq -2) = P(Z \geq 2)$$

This gives us

$$\begin{aligned} P(Z \leq -2) &= P(Z \geq 2) = 1 - P(Z < 2) = 1 - P(Z \leq 2) \\ &\qquad \text{(Point probability is 0, so adding equality sign doesn't matter)} \\ &= 1 - 0.9772 = 0.0228 \end{aligned}$$

1.9	0.9713	0.9719	0.9726	0.9732	0.97
2.0	0.9772	0.9778	0.9783	0.9788	0.97
2.1	0.9821	0.9826	0.9830	0.9834	0.98
2.2	0.9861	0.9864	0.9868	0.9871	0.98

(d) What proportion pay interest within one standard deviation of mean? Within two standard deviations? Within three? Do your answers line up with what you'd expect based on the Empirical Rule?

Since the mean is 27 and the standard deviation is 6, the outcomes within one standard deviation of the mean are $[27-6, 27+6] = [21, 33]$. More formally:

$$P(21 \leq X \leq 33) = P\left(\frac{21-27}{6} \leq Z \leq \frac{33-27}{6}\right) = P(-1 \leq Z \leq 1)$$

The easiest way to proceed using the provided table is to calculate $P(Z > 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587$ and notice that $P(Z \leq -1)$ is equal to the same thing. Thus, due to symmetry, we have

$$P(-1 \leq Z \leq 1) = 1 - 2 \times P(Z \leq -1) = 1 - 2 \times 0.1587 = 0.6826 \approx 0.68$$

Similarly, for within two standard deviations:

$$P(-2 \leq Z \leq 2) = 1 - 2 \times P(Z \leq -2) = 1 - 2 \times 0.0228 = 0.9544 \approx 0.95$$

And within three standard deviations:

$$P(-3 \leq Z \leq 3) = 1 - 2 \times P(Z \leq -3) = 1 - 2 \times 0.0013 = 0.9974 \approx 0.997$$

The results line up exactly with the empirical rule. In fact, the empirical rule is derived using the normal distribution. Thus,

- When working with a normal distribution, the empirical rule is exactly true
- When working with other bell-shaped distributions, the empirical rule gives an approximation

3. Answer the following questions.

(a) Let $Z \sim N(0, 1)$. If $P(Z \leq A) = 0.75$, then what is A?

Doing an inverse look-up of 0.75 on our probability table, we see that $A = 0.675$. (Since the answer lies between two cells on the table, an answer of 0.67 or 0.68 is also fine.)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549

(b) Let $X \sim N(3, 49)$. If $P(X > D) = 0.25$, then what is D?

Standardizing and using the complement rule, we can write this as

$$P(X > D) = P\left(Z > \frac{D-3}{7}\right) = 1 - P\left(Z \leq \frac{D-3}{7}\right) = 0.25$$

This implies that

$$P\left(Z \leq \frac{D-3}{7}\right) = 0.75$$

Since an inverse lookup of 0.75 probability yields $z = 0.675$, it must be that

$$\frac{D-3}{7} = 0.675 \quad \Rightarrow \quad D = 7.725$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549

(c) Let $Y \sim N(\mu, 49)$. If $P(Y < 3) = 0.75$, then what is μ?

Standardizing, we know that:

$$P(Y < 3) = P(Y \leq 3) = P\left(Z \leq \frac{3-\mu}{7}\right) = 0.75$$

Since an inverse lookup of 0.75 probability yields $z = 0.675$, we have

$$\frac{3-\mu}{7} = 0.675 \quad \Rightarrow \quad \mu = -1.725$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549

(d) Let $M \sim N(3, \sigma^2)$. If $P(M > 4) = 0.4$, then what is σ?

Standardizing and using the complement rule, we know that:

$$P(M > 4) = P\left(Z > \frac{4-3}{\sigma}\right) = 1 - P\left(Z \leq \frac{4-3}{\sigma}\right) = 0.4$$

This means that

$$P\left(Z \leq \frac{4-3}{\sigma}\right) = 0.6$$

Since an inverse lookup of 0.6 probability yields $z = 0.255$, it must be that

$$\frac{4-3}{\sigma} = 0.255 \quad \Rightarrow \quad \sigma = 3.9216$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517

4. A cancer drug is produced in one kilogram batches, with an average of six batches produced every hour. Assume the waiting time until the next batch is exponentially distributed.

   (a) What is the probability the production process requires more than 15 minutes to produce the next batch of drugs?

   Denote continuous random variable X as the waiting time until the next batch. From the question, we have that

$$X \sim \text{exponential}(\lambda)$$

   where λ is the success arrival rate (i.e. the rate at which a batch is produced).
   To find what λ is, we are going to utilize the feature of an expoentially distributed variable that

$$E(X) = \frac{1}{\lambda}$$

   So if we can find $E(X)$, which is the average wait time until the next batch, then we can find λ.
   Since the question states that, on average, six batches are produced every hour, this tells us that the average wait time is

$$E(X) = \frac{60}{6} = 10$$

   where the 60 represents 60 minutes are in an hour. Here, we need the average wait time to be in terms of minutes because the question asks probability of waiting more in minute unit.
   Now, with $E(X) = 10$, and that $E(X) = \frac{1}{\lambda}$, we have

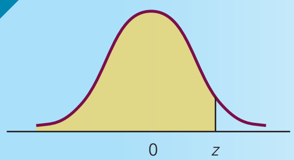
$$\frac{1}{\lambda} = 10 \quad \Rightarrow \quad \lambda = \frac{1}{10}$$

   Finally, since the question asks the probability the production process requires more than 15 minutes to produce the next batch of drugs, we need to solve $P(X > 15)$. By exponential formula:

$$
\begin{aligned}
P(X > 15) &= 1 - P(X \leq 15) = 1 - F(15) = 1 - \left(1 - e^{-\lambda \times 15}\right) \\
&= e^{-\frac{1}{10} \times 15} = 0.2231
\end{aligned}
$$

The probability the production process requires more than 15 minutes to produce the next batch of drugs is 0.2231.

TABLE **3** ***(Continued)***



$P(-\infty < Z < z)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	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