STAT 509: Statistics for Engineers

Chapter 4: Continuous Random Variables and Probability Distributions

Dr. Dewei Wang
Associate Professor
Department of Statistics
University of South Carolina
deweiwang@stat.sc.edu

Chapter 4: Continuous Random Variables and Probability Distributions

Learning Objectives:

1. Determine probabilities from probability density functions.
2. Determine probabilities from cumulative distribution functions, and cumulative distribution functions from probability density functions, and the reverse.
3. Calculate means and variances for continuous random variables.
4. Understand the assumptions for continuous probability distributions.
5. Select an appropriate continuous probability distribution to calculate probabilities for specific applications.
6. Calculate probabilities, means and variances for continuous probability distributions.
7. Standardize normal random variables.

Continuous Random Variables

A continuous random variable is one which takes values in an uncountable set.

They are used to measure physical characteristics such as height, weight, time, volume, position, etc...

Examples

1. Let Y be the height of a person (a real number).
2. Let X be the volume of juice in a can.
3. Let Y be the waiting time until the next person arrives at the server.

Probability Density Functions

Because the number of possible values of X is uncountably infinite, X has a distinctly different distribution from the discrete random variables studied previously.

How to assign probabilities?

- Discrete: probability mass function (pmf)
- Continuous: probability density function (pdf)

Probability Density Functions

Density functions are commonly used in engineering to describe physical systems; e.g., the density of a loading on a long, thin beam.

- For any point x along the beam, the density can be described by a function (in grams/cm).
- Intervals with large loadings correspond to large values for the function. The total loading between points a and b is determined as the integral of the density function from a to b.
- This integral is the area under the density function over this interval, and it can be loosely interpreted as the sum of all the loadings over this interval.

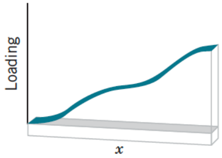


FIGURE 4.1

Density function of a loading on a long, thin beam.

Probability Density Functions

Similarly, a probability density function $f(x)$ can be used to describe the probability distribution of a continuous random variable X.

- If an interval is likely to contain a value for X, its probability is large and it corresponds to large values for $f(x)$.
- The probability that X is between a and b is determined as the integral of $f(x)$ from a to b.

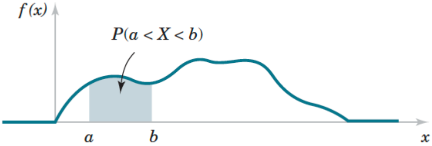


FIGURE 4.2

Probability determined from the area under $f(x)$.

Probability Density Functions

For a continuous random variable X, a **probability density function** is a function such that

(1) $f(x) \geq 0$

(2) $\int_{-\infty}^{+\infty} f(x)dx = 1$

(3) $P(a \leq X \leq b) = \int_a^b f(x)dx =$
area under $f(x)$ from a to b for any a and b

Note that, for a continuous random variable X and *any* value x,

$$P(X = x) = 0.$$

Interpretation: e.g., for the density function of a loading on a long, thin beam, because every point has zero width, the loading at any point is zero. Consequently,

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b).$$

Probability Density Functions

Review of integral:

- polynomial:

$$\int_a^b x^r dx = \frac{x^{r+1}}{r+1}\bigg|_a^b = \frac{b^{r+1}}{r+1} - \frac{a^{r+1}}{r+1}$$

  for $r \neq -1$.

$$\int_a^b \frac{1}{x} dx = \log x \bigg|_a^b = \log b - \log a.$$

- exponential:

$$\int_a^b e^{-x} = \{-e^{-x}\}\bigg|_a^b = e^{-a} - e^{-b}.$$

- $\int_a^b xe^{-x}dx = \int_a^b (-x)de^{-x} = (-x)e^{-x}\big|_a^b - \int_a^b (-e^{-x})dx$

Probability Density Functions

Example 4.1

Let the continuous random variable X denote the current measured in a thin copper wire in milliamperes. Assume that the range of X is $[4.9, 5.1]$mA, and assume that the probability density function of X is $f(x) = 5$ for $4.9 \leq x \leq 5.1$. What is the probability that a current measurement is less than 5 milliamperes?

A common rule: assume that $f(x) = 0$ wherever it is not specifically defined; i.e., $f(x) = 0$ if $x > 5.1$ or $x < 4.9$. Then

$$\begin{aligned} P(X < 5) = P(4.9 \leq X < 5) &= \int_{4.9}^{5} f(x)dx \\ &= \int_{4.9}^{5} 5dx = (5x)\Big|_{4.9}^{5} \\ &= 0.5 \end{aligned}$$

Example 4.2

Let the continuous random variable X denote the diameter of a hole drilled in a sheet metal component. The target diameter is 12.5 millimeters. Most random disturbances to the process result in larger diameters. Historical data show that the distribution of X can be modeled by a probability density function $f(x) = 20e^{-20(x-12.5)}$, for $x \geq 12.5$. If a part with a diameter greater than 12.60 mm is scrapped, what proportion of parts is scrapped?

$$
\begin{aligned}
P(X > 12.60) &= \int_{12.6}^{\infty} f(x)dx = \int_{12.6}^{\infty} 20e^{-20(x-12.5)}dx \\
&= -e^{-20(x-12.5)}\Big|_{12.6}^{\infty} = e^{-20(12.6-12.5)} - 0 \\
&= 0.1353.
\end{aligned}
$$

```
integrand=function(x){20*exp(-20*(x-12.5))}
integrate(integrand,12.6,Inf)
0.1353353 with absolute error < 3.4e-05
```

Example 4.2, continued

What proportion of parts is between 12.5 and 12.6 millimeters?

$$P(12.5 < X < 12.60) = \int_{12.5}^{12.6} f(x)dx = \int_{12.5}^{12.6} 20e^{-20(x-12.5)}dx$$
$$= -e^{-20(x-12.5)}\Big|_{12.5}^{12.6} = 0.8647.$$

In fact, $P(12.5 < X < 12.6) = 1 - P(X > 12.6) = 1 - 0.1353 = 0.8647$.

```
integrand=function(x){20*exp(-20*(x-12.5))}
integrate(integrand,12.5,12.6)
0.8646647 with absolute error < 9.6e-15
```

Cumulative Distribution Functions

An alternative method to describe the distribution of a discrete random variable can also be used for continuous random variables.

Cumulative Distribution Function

The **cumulative distribution function** (cdf) of a continuous random variable X is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

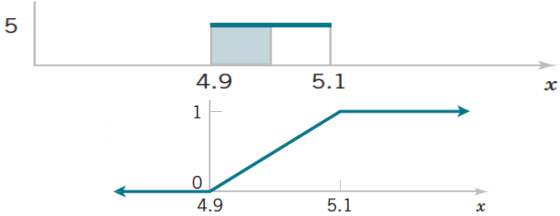
for $-\infty < x < \infty$.

The cumulative distribution function is defined for all real numbers.

Cumulative Distribution Functions

Example 4.1, continued

For the copper current measurement in Example 4.1, Find the cumulative distribution function of the random variable X.



If $x < 4.9$, no way to get $X \leq x$, then $F(x) = 0$ for $x < 4.9$.
If $x > 5.1$, X is definitely less than x, then $F(x) = 1$ for $x > 5.1$.
If $4.9 \leq x \leq 5.1$, then $F(x) = \int_{4.9}^{x} f(u)du = 5(x - 4.9)$.

$$F(x) = \begin{cases} 0 & x < 4.9 \\ 5x - 24.5 & 4.9 \leq x \leq 5.1 \\ 1 & x > 5.1 \end{cases}$$

pdf from cdf

Probability Density Function from the Cumulative Distribution Function

Given $F(x)$

$$f(x) = \frac{dF(x)}{dx}$$

as long as the derivative exists.

Example 4.4, Reaction Time

The time until a chemical reaction is complete (in milliseconds) is approximated by the cumulative distribution function

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-0.01x} & x \geq 0 \end{cases}$$

Determine the probability density function of X. What proportion of reactions is complete within 200 milliseconds?

$$f(x) = \begin{cases} 0 & x < 0 \\ 0.01e^{-0.01x} & x \geq 0 \end{cases}$$

and the probability is $P(X \leq 200) = F(200) = 1 - e^{-2} = 0.8647$.

Mean and Variance of a Continuous Random Variable

Suppose that X is a continuous random variable with probability density function $f(x)$. The mean or expected value of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

The variance of X, denoted as $V(X)$ or σ^2, is

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2.$$

The standard deviation of X is $\sigma = \sqrt{\sigma^2}$.

For any function of X, say $Y = h(X)$, then

$$E(Y) = E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x) dx.$$

Special case: $h(X) = aX + b$: $E[h(X)] = aE(X) + b$ and $V[h(X)] = a^2 V(X)$.

Example 4.1, continued

$$E(X) = \int_{4.9}^{5.1} xf(x)dx = \int_{4.9}^{5.1} 5xdx = \frac{5x^2}{2}\Big|_{4.9}^{5.1} = 5$$

$$\begin{aligned} V(X) &= \int_{4.9}^{5.1} x^2 f(x)dx - 5^2 = \int_{4.9}^{5.1} 5x^2 dx - 25 \\ &= \frac{5x^3}{3}\Big|_{4.9}^{5.1} - 25 \\ &= 0.0033. \end{aligned}$$

```
integrand=function(x){5*x}
integrate(integrand,4.9,5.1)
5 with absolute error < 5.6e-14

integrand=function(x){5*x^2}
integrate(integrand,4.9,5.1)
25.00333 with absolute error < 2.8e-13
```

Example 4.1, continued

X is the current measured in milliamperes. What is the expected value of power when the resistance is 100 ohms?

Solution: We have $P = 10^{-6}RI^2$ where I is the current in milliamperes and R is the resistance in ohms. Thus we define $Y = 10^{-6}100X^2$ where $h(X) = 10^{-6}100X^2$.

$$E(Y) = E[h(X)] = 10^{-4}\int_{4.9}^{5.1} 5x^2 dx = 0.0025 \text{ watts.}$$

```
integrand=function(x){5*x^2*10^(-4)}
integrate(integrand,4.9,5.1)
0.002500333 with absolute error < 2.8e-17
```

Continuous **Uniform** Distribution $X \sim U(a, b)$

- Keyword: Uniform
- pdf: $f(x) = 1/(b-a)$ for $a \leq x \leq b$.
- Mean and variance:

$$\mu = E(X) = (a+b)/2, \quad \sigma^2 = V(X) = (b-a)^2/12.$$

- cdf:

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

StatEngine: Continuous **Uniform** Distribution $X \sim U(a, b)$

```
# Find Mean and variance / plot the pdf/cdf:
    uniform.summary(a,b,plotpdf=c("TRUE","FALSE"),
                    plotcdf=c("TRUE","FALSE"))
# Find P(lb<X<ub):
    uniform.prob(a,b,lb,ub)
# Find x such that P(X<x)=q for a given q:
    uniform.quantile(a,b,q)
```

Example: In Example 4.1, the random variable X has a continuous uniform distribution on $[4.9, 5.1]$. The probability density function of X is $f(x) = 5, 4.9 \leq x \leq 5.1$. Find the probaiblity $P(4.95 < X < 5)$ and the value x such that $P(X > x) = 0.1$

```
    a=4.9;b=5.1;uniform.summary(a,b)
    uniform.prob(a,b,4.95,5.0)
    0.25
    uniform.quantile(a,b,1-0.1)
    5.08
```

Normal Distribution $X \sim N(\mu, \sigma^2)$

- Keyword: Normal. Notation: $X \sim N(\mu, \sigma^2)$.
- pdf:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty.$$

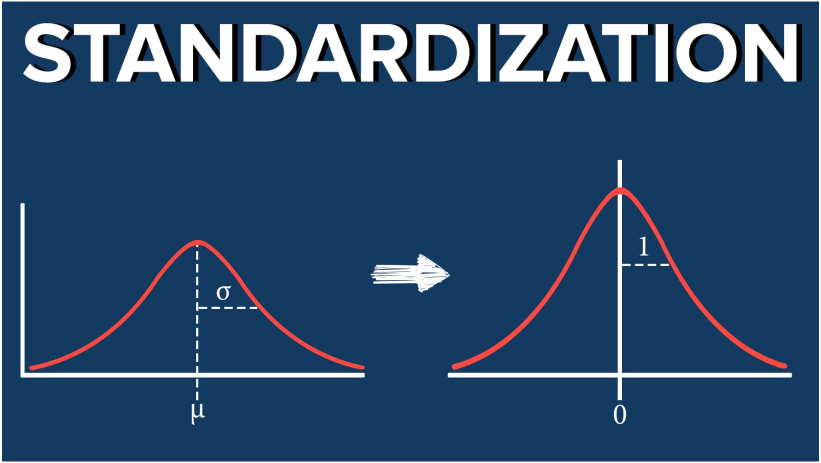
- Mean and variance: $E(X) = \mu, \quad V(X) = \sigma^2$.
- Standardization: If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1); \textbf{standard normal distribution}$$

- cdf: $F(x) = \Phi(\frac{x-\mu}{\sigma})$ has no closed-form expression, where $\Phi(x)$ is the cdf of $N(0,1)$.

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx$$

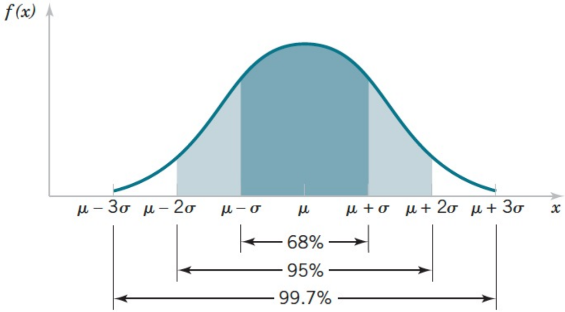
Normal Distribution $X \sim N(\mu, \sigma^2)$



Normal Distribution $X \sim N(\mu, \sigma^2)$

Empirical Rule

$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$
$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$
$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$



StatEngine: **Normal** Distribution $X \sim N(\mu, \sigma^2)$

```
# Find Mean and variance / plot the pdf/cdf:
    normal.summary(mu,sigma,plotpdf=c("TRUE","FALSE"),
                    plotcdf=c("TRUE","FALSE"))

# Find P(lb<X<ub):
    normal.prob(mu,sigma,lb,ub)

# Find x such that P(X<x)=q for a given q:
    normal.quantile(mu,sigma,q)
```

Normal Distribution: Example

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with $\mu = 10$ and $\sigma = 2$ mA.

1. What is the probability that the current measurement is between 9 and 11 mA?

   Solution: Let X be the current measurement.

   ```
   P(9<X<11)=normal.prob(10,2,9,11)=0.3829
   ```

2. Determine the value for which the probability that a current measurement is below 0.98.

   Solution: Let X be the current measurement. We need to find x such that $P(X < x) = 0.98$

   ```
   x=normal.quantile(10,2,0.98)=14.1075
   ```

Normal Distribution: Example, continued

3. Suppose the value μ can be adjusted while $\sigma = 2$mA is fixed. At which μ, we can have $P(X > 13) = 0.2$?

   Solution: $X \sim N(\mu, 2^2)$ implies $Z = (X - \mu)/2 \sim N(0, 1)$. Thus

   $$0.2 = P(X > 13) = P\left(\frac{X-\mu}{2} > \frac{13-\mu}{2}\right) = P\left(Z > \frac{13-\mu}{2}\right),$$

   Thus

   $\frac{13-\mu}{2}$`=normal.quantile(0,1,1-0.2)=0.8416`

   which leads to

   $$\mu = 11.3168.$$

4. Now suppose $\mu = 10$mA is fixed while σ can be adjusted. At which σ, we can have $P(X > 14) = 0.1$? (Answer:3.1212)

Exponential Distribution $X \sim \mathrm{Exp}(\lambda)$

The random variable X that equals the distance between successive events from a Poisson process with mean number of events $\lambda > 0$ per unit interval is an exponential random variable with parameter λ.

- Keyword: Exponential, Poisson process. Notation: $X \sim \mathrm{Exp}(\lambda)$.
- Usage: model time (time to next event, lifetime of a product).
- pdf:
$$f(x) = \lambda e^{-\lambda x}, \quad 0 < x < \infty.$$
- Mean and variance: $\mu = E(X) = \lambda^{-1}, \quad \sigma^2 = V(X) = \lambda^{-2}.$
- cdf: $F(x) = 1 - e^{-\lambda x}, \quad 0 < x < \infty.$
- Lack of Memory Property:
$$P(X < t_1 + t_2 | X > t_1) = P(X < t_2).$$

StatEngine: **Exponential** Distribution $X \sim \mathrm{Exp}(\lambda)$

```
# Find Mean and variance / plot the pdf/cdf:
    exponential.summary(lambda,plotpdf=c("TRUE","FALSE"),
                        plotcdf=c("TRUE","FALSE"))

# Find P(lb<X<ub):
    exponential.prob(lambda,lb,ub)

# Find x such that P(X<x)=q for a given q:
    exponential.quantile(lambda,q)
```

Exponential Distribution: Example

In a large corporate computer network, user log-ons to the system can be modeled as a Poisson process with a mean of 25 log-ons per hour.

1. What is the probability that there are no log-ons in an interval of 6 minutes?

   Solution: Let X denote the time in hours from the start of the interval until the first log-on. Then $X \sim \text{Exp}(\lambda = 25)$. Known 6 minuets=0.1 hour!

   `P(X>0.1)=exponential.prob(25,0.1,Inf)=0.0821`
   `or =1-P(0<X<0.1)=1-exponential.prob(25,0,0.1)`

2. Determine the interval of time such that the probability that no log-on occurs in the interval is 0.90.

   Solution: Find x such that $P(X > x) = 0.9$

   `x=exponential.quantile(25,1-0.9)=0.0042`

Exponential Distribution: Example (lack of memory)

Let X denote the time between detections of a particle with a Geiger counter and assume that X has an exponential distribution with $E(X) = 1.4$ minutes (i.e., $\lambda = 1/1.4$). The probability that we detect a particle within 0.5 minute of starting the counter is

$$P(X < 0.5) = F(0.5) = 1 - e^{-0.51.4} = 0.30.$$

Now, suppose that we turn on the Geiger counter and wait 3 minutes without detecting a particle. What is the probability that a particle is detected in the next 30 seconds?

Solution: It asks for the conditional property $P(X < 3.5|X > 3)$. Because we have already been waiting for 3 minutes, we feel that a detection is "due." That is, the probability of a detection in the next 30 seconds ($= 0.5$ minuets) should be higher than 0.3. However, for an exponential distribution, this is not true because of the lack of memory property. The fact that we have waited 3 minutes without a detection does not change the probability of a detection in the next 30 seconds. We have $P(X < 3.5|X > 3) = P(X < 0.5) = 0.3$.

Gamma Distribution $X \sim \text{Gamma}(r, \lambda)$

- Keyword: Gamma, Poisson process. Notation: $X \sim \text{Gamma}(r, \lambda)$. r: shape, λ: scale.
- Usage: model time (time to next event, lifetime of a product).
- pdf:

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, \quad 0 < x < \infty,$$

  where $\Gamma(r)$ is the gamma function $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$ for $r > 0$. It can shown that $\Gamma(r) = (r-1)\Gamma(r-1)$ and $\Gamma(r) = (r-1)!$, $\Gamma(1) = 0! = 1$, $\Gamma(0.5) = \sqrt{\pi}$.
- Mean and variance: $\mu = E(X) = r/\lambda, \quad \sigma^2 = V(X) = r/\lambda^2.$
- cdf: $F(x)$ no closed-form expression.

When r is an integer, it is also called an **Erlang** distribution. X can be interpreted as the distance until the next rth event in a Poisson process with mean number of events $\lambda > 0$ per unit interval.

StatEngine: **Gamma** Distribution $X \sim \text{Gamma}(r, \lambda)$

```
# Find Mean and variance / plot the pdf/cdf:
    gamma.summary(r,lambda,plotpdf=c("TRUE","FALSE"),
                    plotcdf=c("TRUE","FALSE"))

# Find P(lb<X<ub):
    gamma.prob(r,lambda,lb,ub)

# Find x such that P(X<x)=q for a given q:
    gamma.quantile(r,lambda,q)
```

Gamma Distribution: Example

The time to prepare a slide for high-throughput genomics is a Poisson process with a mean of two hours per slide.

1. What is the probability that 10 slides require more than 25 hours to prepare?

   Solution: Let X denote the time to prepare 10 slides. Because of the assumption of a Poisson process, $X \sim \text{Gamma}(r = 10, \lambda = 1/2)$.

   `P(X>25)=gamma.prob(10,0.5,25,Inf)=0.2014`

2. What are the mean and standard deviation of the time to prepare 10 slides?

   `gamma.summary(10,0.5) #:` $\mu = 20, \ \sigma = 6.3246.$

3. The slides will be completed by what length of time with probability equal to 0.95?

   Solution: Find x such that $P(X \leq x) = 0.95$.

   `x=gamma.quantile(10,0.5,0.95)=31.4104`

Weibull Distribution $X \sim \text{Weibull}(\beta, \delta)$

- Keyword: Weibull. Notation: $X \sim \text{Weibull}(\beta, \delta)$. $\beta > 0$: shape, $\delta > 0$: scale.
- Usage: model time (time to next event, lifetime of a product).
- pdf:

$$f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\delta}\right)^{\beta}\right], \quad 0 < x < \infty,$$

- Mean and variance:

$$E(X) = \delta\Gamma(1+\beta^{-1}), \quad V(X) = \delta^2\Gamma(1+2/\beta) - \delta^2[\Gamma(1+1/\beta)]^2.$$

- cdf:

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\delta}\right)^{\beta}\right], \quad 0 < x < \infty.$$

StatEngine: **Weibull** Distribution $X \sim \text{Weibull}(\beta, \delta)$

```
# Find Mean and variance / plot the pdf/cdf:
    weibull.summary(beta,delta,plotpdf=c("TRUE","FALSE"),
                    plotcdf=c("TRUE","FALSE"))

# Find P(lb<X<ub):
    weibull.prob(beta,delta,lb,ub)

# Find x such that P(X<x)=q for a given q:
    weibull.quantile(beta,delta,q)
```

Weibull Distribution: Example

The time to failure (in hours) of a bearing in a mechanical shaft is satisfactorily modeled as a Weibull random variable with $\beta = 2$ and $\delta = 5000$ hours.

1. Determine the mean and standard deviation of the time until failure.

   Solution: Let X denote the time to failure.
   $X \sim \text{Weibull}(\beta = 2, \delta = 5000)$.

   `weibull.summary(2,5000) #:` $\mu = 4431.135, \;\; \sigma = 2316.257$

2. Determine the probability that a bearing lasts at least 6000 hours.

   `P(X>=6000)=weibull.prob(2,5000,6000,Inf)=0.2369`

3. Find x such that $P(X > x) = 0.05$.

   `x=weibull.quantile(2,5000,1-0.05)=8654.092`

Lognormal Distribution $X \sim \text{logN}(\theta, \omega^2)$

Let $W \sim N(\theta, \omega^2)$, then $X = \exp(W)$ is a **lognormal** random variable; that is, $\log X$ is a normal random varialbe.

- Keyword: Lognormal. Notation: $X \sim \text{logN}(\theta, \omega^2)$ means $\log X \sim N(\theta, \omega^2)$. θ: mean of $\log X$. ω^2: variance of $\log X$.
- Usage: model time (time to next event, lifetime of a product).
- pdf:

$$f(x) = \frac{1}{x\omega\sqrt{2\pi}} \exp\left[-\frac{(\log x - \theta)^2}{2\omega^2}\right], \quad 0 < x < \infty,$$

- Mean and variance:

$$E(X) = \exp\{\theta + \omega^2/2\}, \quad V(X) = e^{2\theta+\omega^2}(e^{\omega^2} - 1).$$

- cdf: no closed form.

StatEngine: **Lognormal** Distribution $X \sim \text{logN}(\theta, \omega^2)$

```
# Find Mean and variance / plot the pdf/cdf:
    lognormal.summary(theta,omega,plotpdf=c("TRUE","FALSE"),
                    plotcdf=c("TRUE","FALSE"))

# Find P(lb<X<ub):
    lognormal.prob(theta,omega,lb,ub)

# Find x such that P(X<x)=q for a given q:
    lognormal.quantile(theta,omega,q)
```

Lognormal Distribution: Example

The lifetime (in hours) of a semiconductor laser has a lognormal distribution with $\theta = 10$ and $\omega = 1.5$.

1. Determine the mean and standard deviation of the lifetime.

   Solution: Let X denote the lifetime.
   $X \sim \text{LogN}(\theta = 10, \omega = 1.5)$.

   `lognormal.summary(10,1.5) #:` $\mu = 67846.29, \quad \sigma = 197661.5$
2. What is the probability that the lifetime exceeds $10,000$ hours?

   `P(X>10000)=lognormal.prob(10,1.5,10000,Inf)=0.7007`
3. What lifetime is exceeded by 99% of lasers?

   Solution: Find x such that $P(X > x) = 0.99$.

   `x=lognormal.quantile(10,1.5,1-0.99)= 672.1478`

Beta Distribution $X \sim \text{Beta}(\alpha, \beta)$

- Keyword: Beta. Notation: $X \sim \text{Beta}(\alpha, \beta)$. Both $\alpha > 0$ and $\beta > 0$ are shape parameters.
- Usage: model proportion.
- pdf:

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1,$$

- Mean and variance:

$$E(X) = \frac{\alpha}{\alpha+\beta}, \quad V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

- cdf: no closed form.

StatEngine: **Beta** Distribution $X \sim \text{Beta}(\alpha, \beta)$

```
# Find Mean and variance / plot the pdf/cdf:
    beta.summary(alpha,beta,plotpdf=c("TRUE","FALSE"),
                    plotcdf=c("TRUE","FALSE"))

# Find P(lb<X<ub):
    beta.prob(alpha,beta,lb,ub)

# Find x such that P(X<x)=q for a given q:
    beta.quantile(alpha,beta,q)
```

Beta Distribution: Example

The service of a constant-velocity joint in an automobile requires disassembly, boot replacement, and assembly. Suppose that the proportion of the total service time for disassembly follows a beta distribution with $\alpha = 2.5$ and $\beta = 1$.

1. Determine the mean and standard deviation of the proportion.

   Solution: Let X denote the proportion of service time for disassembly. $X \sim \text{Beta}(\alpha = 2.5, \beta = 1)$.

   `beta.summary(2.5,1) #:` $\mu = 0.7143, \ \ \sigma = 0.2130$

2. What is the probability that a disassembly proportion exceeds 0.7?

   `P(X>0.7)=beta.prob(2.5,1,0.7,Inf)=0.59`

3. Find x such that $P(X \leq x) = 0.99$.

   `x=beta.quantile(2.5,1,0.99)=0.996`