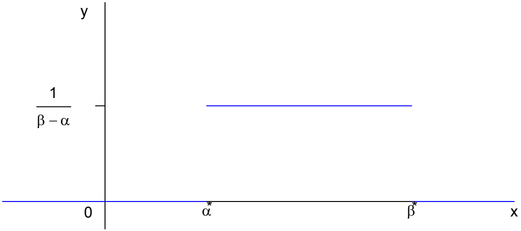
Statistics ST2334 Topic 5: Probability Densities (part b)

2011/2012 Semester 2

Uniform Distribution

- The probability density function for the uniform distribution with parameters α, β is

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha}, & \text{if } \alpha \le x \le \beta \\ 0, & \text{otherwise} \end{cases}$$



- All values within the interval $[\alpha, \beta]$ are "equally likely".

Uniform Distribution

- Expectation:

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} x\, dx = \frac{1}{\beta - \alpha} \frac{x^2}{2} \Big|_{\alpha}^{\beta} = \frac{\alpha + \beta}{2}.$$

- Variance:

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\, dx = \frac{1}{12} (\beta - \alpha)^2.$$

- CDF:

$$F(x) = \int_{-\infty}^{x} f(t) dt = \begin{cases} 0, & \text{if } x < \alpha \\ \frac{x - \alpha}{\beta - \alpha}, & \text{if } \alpha \leq x \leq \beta \\ 1, & \text{if } x > \beta \end{cases}$$

Example: Taking Bus to Class

- A student attending ST2334 travels to LT27 by bus 95, which arrives at the bus stop near his home every 15 min. The bus ride takes 20 min to get to LT27.
- Let X be the total time taken from his bus stop. We assume $X \sim Uniform(20, 35)$
- How long on average does he take to get to LT27?

  The mean time he takes is

$$E(X) = (\alpha + \beta)/2 = (20 + 35)/2 = 27.5$$

Example: Taking Bus to Class

- If he arrives at the bus stop at 9:35am, what is the chance that he arrives at LT27 before 10:00am?

$$
\begin{aligned}
P(\text{arrive on time}) &= P(X \leq 25) \\
&= \frac{25-20}{35-20} \\
&= \frac{1}{3}
\end{aligned}
$$

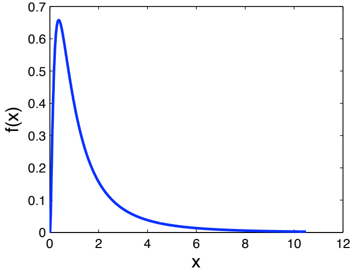
Log-Normal Distribution

- X follows a log-normal distribution with parameters (α, β), if $\ln(X)$ follows a normal distribution $N(\alpha, \beta^2)$
- Its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\beta^2}} x^{-1} e^{-\frac{(\ln x - \alpha)^2}{2\beta^2}}, & \text{if } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

Log-Normal Distribution

- For $\alpha = 0$ and $\beta = 1$,



- Expectation:

$$E(X) = e^{\alpha+\beta^2/2}$$

- Variance:

$$Var(X) = e^{2\alpha+\beta^2}(e^{\beta^2} - 1)$$

Log-Normal Distribution

- Let X have log-normal distribution with parameters α, β. Then, the cdf is given by

$$
\begin{aligned}
F(x) = P(X \leq x) &= P(\ln X \leq \ln x) \\
&= P\left(\frac{\ln X - \alpha}{\beta} \leq \frac{\ln x - \alpha}{\beta}\right) \\
&= P\left(Z \leq \frac{\ln x - \alpha}{\beta}\right) \\
&= \Phi\left(\frac{\ln x - \alpha}{\beta}\right)
\end{aligned}
$$

- Consequently,

$$P(a \leq X \leq b) = \Phi\left(\frac{\ln b - \alpha}{\beta}\right) - \Phi\left(\frac{\ln a - \alpha}{\beta}\right)$$

Example: Black-Scholes Model

Under the Black-Scholes Model, changes in the logarithm of stock market indices are normal. Historically, the daily changes in the logarithm of STI has mean 0.000212 and variance 0.000168. What is the probability that STI is up more than 2% on any given day?

- Let S_i be the STI on day i. We are interested in $P(S_{i+1}/S_i > 1.02)$.
- $\ln(S_{i+1}/S_i) = \ln(S_{i+1}) - log(S_i) \sim N(0.000212, 0.0130^2)$, S_{i+1}/S_i has log-normal distribution with $\alpha = 0.000212, \beta = 0.0130$

$$\begin{aligned}
P(S_{i+1}/S_i > 1.02) &= 1 - P(S_{i+1}/S_i < 1.02) \\
&= 1 - \Phi\left(\frac{\ln(1.02) - 0.000212}{0.0130}\right) \\
&= 1 - \Phi(1.51) \\
&= 0.0655
\end{aligned}$$

Gamma Distribution

- The probability density function for the Gamma distribution with parameters $\alpha > 0, \beta > 0$ is

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx$.

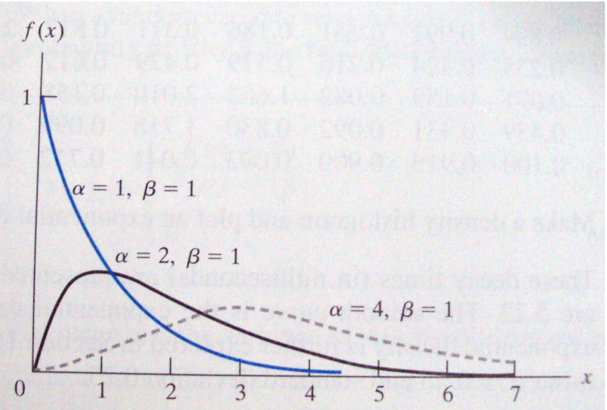
  - Integration by parts shows that for $\alpha > 1$,

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

  - In the special case where α is integer, $\Gamma(\alpha) = (\alpha - 1)!$

Gamma Distribution

- Gamma family of distributions



Gamma Distribution

If $X \sim Gamma(\alpha, \beta)$,

- Expectation
$$E(X) = \alpha\beta$$
- Variance
$$Var(X) = \alpha\beta^2$$
- In general, there is no closed form CDF for X.

Exponential Distribution

- The exponential distribution is an important special case of the gamma distribution.
- It corresponds to the gamma distribution with $\alpha = 1$.
- The probability density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

  - As before, the parameter $\beta > 0$
- and it follows that

$$E(X) = \beta \text{ and } Var(X) = \beta^2$$

.

Exponential Distribution

- The CDF does have a closed form expression given by

$$F(x) = \begin{cases} 1 - e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Let X follow the exponential distribution with parameter β. Note that

$$P(X > x) = 1 - F(x) = e^{-x/\beta}.$$

It follows that for all $s, t > 0$,

$$\begin{aligned} P(X > s + t | X > s) &= P(X > s + t)/P(X > s) \\ &= e^{-(s+t)/\beta}/e^{-s/\beta} \\ &= e^{-t/\beta} = P(X > t) \end{aligned}$$

- This is the memoryless property of the exponential distribution.

Example: Light Bulbs

- The life time of ABC light bulbs follows the exponential distribution. If the average life time of ABC light bulbs is 3 years. What is the probability of a ABC light bulb lasting longer than 4 years?
  Let X be the lifetime of the ABC light bulb.

  $$E(X) = 3 \Rightarrow \beta = 3$$

  Hence,

  $$P(X > 4) = 1 - F(4) = e^{-4/3} = 0.264$$

- What is the probability of the light bulb lasting another 4 years given that it has already lasted 4 years?

Checking if Data is Normal

- On one hand, the distribution of many real data can be well approximated by normal distribution.
- On the other hand, many statistical methods are based on the assumption that the collected data are normally distributed.
- Normality is widely used as basic, preliminary conditions in many statistical methods. But mis-specification can result serious errors in statistical inference.

Methods of Checking Normality

- Plot a histogram of the data: see if histogram is closed to the familiar bell shape. Skewed histogram implies the data are not normally distributed.
  - However, it is some times difficult to judge how close a symmetric bell-shaped like histogram is the normal density.
- Normal scores plot (normal quantile plot, QQ-plot) provides an effective way of checking normality.

Normal Scores

- Normal Scores ($m_i, i = 1, ..., n$): values of z that separate the region under the curve of standard normal density into $n + 1$ parts of equal areas.

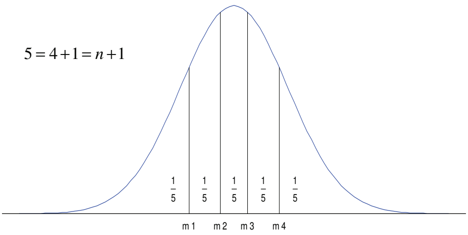
$$z_{1-1/(n+1)}, z_{1-2/(n+1)}, , \cdots, z_{1-n/(n+1)}$$

- These n values separate the region under the curve of standard normal density into n+1 parts of equal areas.
- Normal scores refers to an ideal/"perfect" sample from the standard normal distribution.
- Recall: for any $0 < \alpha < 1$, is the one satisfying $P(Z > z_\alpha) = \alpha$.

Normal Scores Example

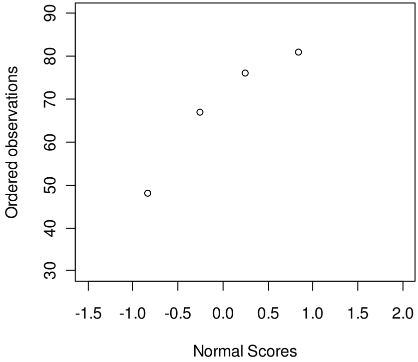
- Example, n = 4, then,

$$\begin{aligned} m_1 &= z_{0.80} = -0.84 \\ m_2 &= z_{0.60} = -0.25 \\ m_3 &= z_{0.40} = 0.25 \\ m_4 &= z_{0.20} = 0.84 \end{aligned}$$



Draw the normal score plot

- We simply sort the data in ascending order, then plot it against the normal scores.
- Toy example: suppose we have four observations 67, 48, 76, 81.
- Sort: 46, 67, 76, 81. Normal Scores: -0.84, -0.25, 0.25, 0.84

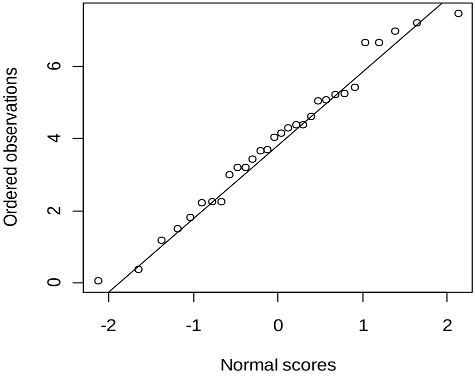


Reading Normal Score Plot

- If the data follow the standard normal distribution, the points of the plot will be roughly on the 45^o line.
- If the data follow a normal distribution, the points of the plot will be roughly follow a line.
- By observing whether points of normal scores plot are roughly on a line, we can conclude whether it is ok to assume normality.
- Note that we usually need a minimum 15 observations (i.e. $n \geq 15$) in order to evaluate its agreement with normality.

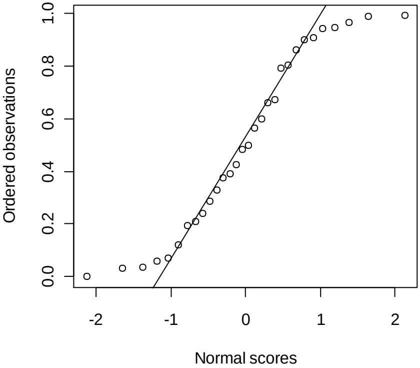
Normal Score Plot Examples

- When the data are normally distributed



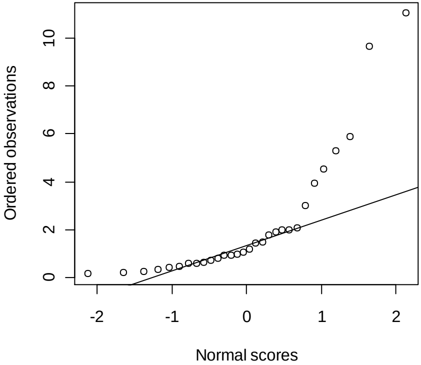
Normal Score Plot Examples

- When the data are uniformly distributed



Normal Score Plot Examples

- When the data are log-normally distributed



Transforming Observations to Near Normality

When the histogram or normal scores plot indicate that the assumption of a normal distribution is invalid, transformations of the data can often improve the agreement with normality.

- To make large values smaller (when the distribution skewed to the right), we can use

$$-\frac{1}{x}, \quad \ln(x), \quad x^{1/4}, \quad x^{1/2}$$

- To make large values larger (when the distribution skewed to the left), we can use

$$x^2, \quad x^3$$

Black-Scholes STI?

- R code will be posted.