STAT 511

Lecture 10: Other Continuous Distributions and Probability Plots

Devore: Section 4.4-4.6

Prof. Michael Levine

February 12, 2019

Gamma Distribution

- **Gamma function** is a natural extension of the factorial
- For any $\alpha > 0$,

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

- Properties:
  1. If $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
  2. $\Gamma(n) = (n-1)!$ for any $n \in Z^+$
  3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

- The natural definition of a density based on the gamma function is

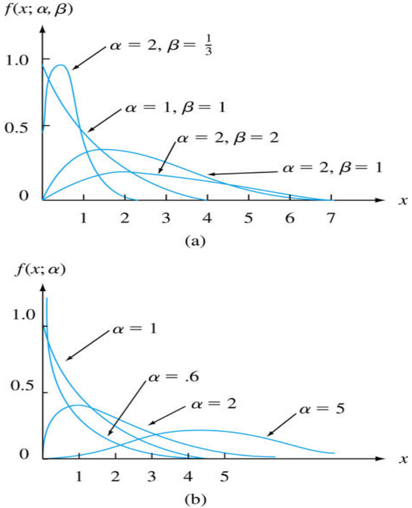
$$f(x;\alpha) = \begin{cases} \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

- A **gamma density** with parameters $\alpha > 0$, $\beta > 0$ is

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1}e^{-x/\beta} \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

- α is a **shape parameter**, β is a **scale parameter**
- The case $\beta = 1$ is called the *standard gamma distribution*

Gamma pdf: graphical illustration



Gamma Distribution Parameters

- The mean and variance of a random variable X having the gamma distribution $f(x;\alpha,\beta)$ are

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

- Let X have a gamma distribution with parameters α and β
- Then $P(X \le x) = F(x;\alpha,\beta) = F(x/\beta;\alpha)$
- In the above,

$$F(x;\alpha) = \int_0^x \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}\,dy$$

is an **incomplete gamma function**. It is defined for any $x > 0$.

Exponential Distribution as a Special Case of Gamma Distribution

- Assume that $\alpha = 1$ and $\beta = \frac{1}{\lambda}$.
- Then,

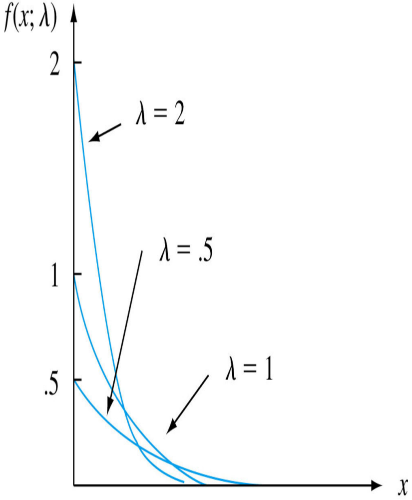
$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

- Its mean and variance are

$$E(X) = \frac{1}{\lambda} \text{ and } V(X) = \frac{1}{\lambda^2}$$

- Note that $\mu = \sigma$ in this case

Exponential pdf: graphical illustration



Exponential cdf

- Exponential cdf can be easily obtained by integrating pdf, unlike the cdf of the general Gamma distribution
- The result is

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

Example I

- Suppose the response time X at an on-line computer terminal has an exponential distribution with expected response time 5 sec
- $E(X) = \frac{1}{\lambda} = 5$ and $\lambda = 0.2$
- The probability that the response time is between 5 sec and 10 sec is

$$P(5 \leq X \leq 10) = F(10; 0.2) - F(5; 0.2) = 0.233$$

Example II

- On average, 3 trucks per hour arrive to a given warehouse to be unloaded. What is the probability that the time between arrivals is less than 5 min? At least 45 min?
- Arrivals follow a Poisson process with parameter $\lambda = 3$; the respective exponential distribution parameter is also $\lambda = 3$
- $$\int_0^{1/12} 3\exp(-3x)\,dx = 1 - \exp(-1/4) = 0.221$$
- $$\int_{3/4}^{\infty} 3\exp(-3x) = \exp(-9/4) = 0.105$$

Weibull Distribution

- X is said to have a Weibull distribution if

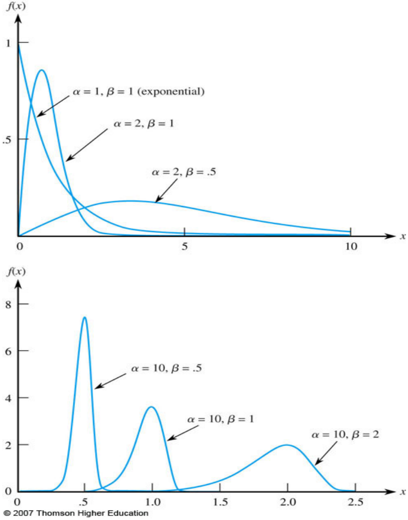
$$f(x;\alpha,\beta)=\begin{cases}\frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}e^{-(x/\beta)^{\alpha}} & \text{if } x\geq 0\\ 0 \text{ otherwise}\end{cases}$$

- Note that Weibull distribution is yet another generalization of exponential; indeed, if $\alpha=1$, Weibull pdf is exponential with $\lambda=\frac{1}{\beta}$.
- The Weibull cdf is

$$F(x;\alpha,\beta)=\begin{cases}1-e^{-(x/\beta)^{\alpha}} & \text{if } x\geq 0\\ 0 \text{ otherwise}\end{cases}$$

- The parameters $\alpha>0$ and $\beta>0$ are referred to as the shape and scale parameters, respectively.

Weibull densities: graphical illustration



Alternative definition of Weibull distribution

- Weibull distribution is used in survival analysis, many processes in engineering, such as engine emissions, degradation data analysis etc
- Often, the Weibull distribution is also defined as bounded from below.
- Let X be the corrosion weight loss for a small square magnesium alloy plate. The plate has been immersed for 7 days in an inhibited 20% solution of $MgBr_2$.
- Suppose the min possible weight loss is $\gamma = 3$ and $X = 3$ has a Weibull distribution with $\alpha = 2$ and $\beta = 4$.
- $$F(x; 2, 4, 3) = \begin{cases} 1 - e^{-[(x-3)/4]^2} \text{ if } x \geq 3 \\ 0 \text{ if } x < 3 \end{cases}$$
- Then, e.g. probability

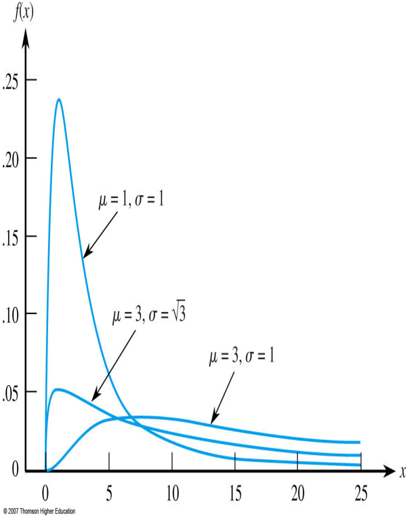
$$P(X > 3.5) = 1 - F(3.5; 2, 4, 3) = .985$$

Lognormal distribution

- If $Y = \log(X)$ is normal, X is said to have a lognormal distribution
- Its pdf is

$$\begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-[\log(x)-\mu]^2/(2\sigma^2)} \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases}$$

Lognormal densities: graphical illustration



- Note that μ and σ^2 are NOT the mean and variance of the lognormal distribution. Those are $E(X) = \exp(\mu + \sigma^2/2)$ and $Var(X) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$. Also note that the lognormal distribution is not symmetric but rather positively skewed.
- Lognormal distribution is commonly used to model various material properties.
- To find probabilities related to the lognormal distribution note that

$$P(X \leq x) = P(\log(X) \leq \log(x))$$
$$= P\left(Z \leq \frac{\log(x) - \mu}{\sigma}\right) = \Phi\left(\frac{\log(x) - \mu}{\sigma}\right)$$

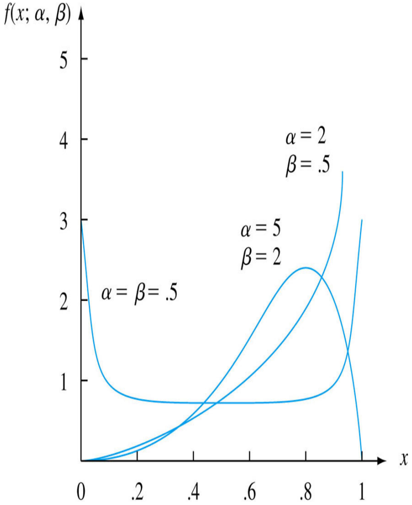
Beta Distribution

- Beta Distribution is used to model random quantities that are positive on a closed interval $[A, B]$ only. It gives us much more flexibility than the uniform distribution.
- A RV X is said to have a beta distribution with parameters A, B, $\alpha > 0$ and $\beta > 0$ if its pdf is

$$f(x; \alpha, \beta, A, B) = \begin{cases} \frac{1}{B-A} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{x-A}{B-A}\right)^{\alpha-1} \left(\frac{B-x}{B-A}\right)^{\beta-1} \\ \text{ if } A \leq x \leq B \\ 0 \text{ if } x < 0 \end{cases}$$

- The mean and variance of X are $\mu = A + (B-A)\frac{\alpha}{\alpha+\beta}$ and $\sigma^2 = \frac{(B-A)^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Beta densities: graphical illustration



Example:Introduction

- Beta distribution is commonly used to model variation in the proportion of a quantity occurring in different samples (e.g. proportion of a 24-hour day that an individual is asleep or the proportion of a certain element in a chemical compound)
- Consider PERT - Program Evaluation and Review Technique. PERT is a method used to coordinate various activities that a large project may consist of. The standard assumption is that the time necessary to complete a particular activity once it has been started is Beta on an interval $[A, B]$ with some α and β.

Example: Calculations

- Suppose that in constructing a single family house the time X (in days) needed to lay a foundation is Beta on $[2, 5]$ with $\alpha = 2$ and $\beta = 3$.
- All Beta-related calculations in R are done using the so-called **standard beta disrtibution** that is defined on $[0, 1]$
- To compute $P(X \leq 3)$ we need to transform our beta distribution to standard one: pbeta((3 - 2)/(5 - 2), shape1 = 2, shape2 = 3)

Sample Percentiles

- For a given p, the solution of the equation $F(\eta(p)) = p$ gives a $100p$th percentile of the continuous distribution described by F.
- Now, order the n-sample observations from smallest to largest as $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$.

Sample percentile definition

- We know how to define the sample median for both odd and even n. For example, if $n = 10$, the median is $\frac{1}{2}(x_{(5)} + x_{(6)})$. In the same way, the lower quartile (25% percentile) can be taken to be $x_{(3)}$ - the median of the lower half. Again, we regard it to be half in the lower group and half in the upper group etc.
- In general, we say that the ith smallest observation in the list is taken to be the $\frac{100(i-0.5)}{n}$ th sample percentile
- The intermediate percentages are obtained by linear interpolation. For example, if $n = 10$, one obtains 5% percentile and 15% percentile by assuming $i = 1$ and $i = 2$ in the above expression. Then, 10th percentile is halfway between the above two.

Normal Probability Plot

- A plot of the n pairs $[\frac{100(i-0.5)}{n}$th z percentile, ith smallest observation] is referred to as the **normal probability plot**.
- Generate 20 observations from the Gamma distribution with $\alpha = 2$ and $\beta = 1$
- The following normal probability plot illustrates just how much unlike normal distribution gamma distribution is: x $<-$ rgamma(20,2) qqnorm (x)

Example of Several Probability Plots for different samples from the Standard Normal Distribution

- If the observations are drawn from $Z \sim N(0,1)$, a straight line through the origin at 45 deg is expected.
- This example helps learning how to avoid overinterpreting deviations from the straight line pattern
- R code:
  1. par(mfrow=c(2,2))
  2. for (I in 1:4){
  3. x < − rnorm (100)
  4. qqnorm(x)
  5. }

What to expect from a normal probability plot?

- If the sample observations are drawn from a normal distribution $N(\mu, \sigma^2)$, the points should fall close to a straight line with slope σ and intercept μ.
- The following are the common deviations from normality:
  - "Lighter tails" than the normal ones but the distribution is symmetric
  - "Heavier tails" than the normal ones but the distribution is symmetric
  - The distribution is skewed

- When the distribution is light-tailed, the extrema are not as "extreme" as what you'd expect from a normal distribution. The result is an S-shaped pattern.
- When the distribution is a heavy-tailed one, the result is also an S-curve but with different orientation. In this case, the extrema are more "extreme" than those of the normal distribution.
- If the underlying observation is positively skewed, the smallest sample observations are larger than what is expected from the normal observations and the same is true for the largest ones. The result is a cup patterns...This is commonly observed when the data is generated by the lognormal distribution.

Data example: forearms' length dataset

- Forearms' lengths are an example of linear measurement
- Can they be normally distributed?
  1. forearms $<-$ as.vector(as.matrix(forearms))
  2. par(mfrow=c(1,2))
  3. boxplot(forearms, main="Boxplot")
  4. qqnorm (forearms)