Topic 7 - Continuous Distributions II

STAT 511

Professor Bruce Craig

Background Reading

Devore : Section 4.1 - 4.6

Gamma Distribution

- Family of pdf's that yields a wide variety of skewed distributions
- Distribution relies on **gamma function**

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x)dx \text{ for } \alpha > 0$$

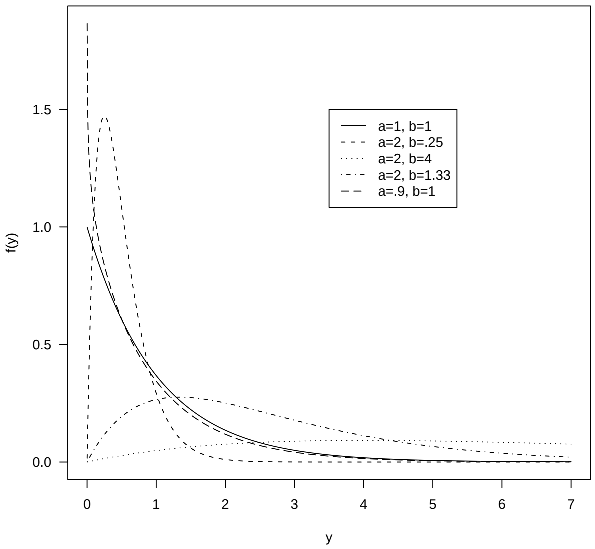
  - For $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
  - For positive integer n, $\Gamma(n) = (n - 1)!$
  - $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- Gamma pdf is

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp(-x/\beta) \text{ for } x \geq 0$$

- Standard Gamma family has $\beta = 1$

Gamma Distribution

- Selection of Gamma Family Densities



Gamma Distribution

- Distribution parameters: α, β
- Can show
  - $E(X) = \alpha\beta$
  - $V(X) = \alpha\beta^2$
- $P(X < x)$ given in Table A.4 for
  - $\beta = 1$
  - $\alpha = 1, 2, ..., 10$
  - $x = 1, 2, ..., 15$
- For other β, $P(X < x) = F(x/\beta, \alpha)$

Example - Problem 57

Time spent on a computer (X) is gamma distributed with mean 20 min and variance 80 min^2.

- What are the values of α and β?

$$\alpha\beta = 20 \text{ and } \alpha\beta^2 = 80$$
$$\beta = 80/20 = 4 \text{ and } \alpha = 20/4 = 5$$

- What is $P(X < 24)$?

  Since $\beta \neq 1$, we need to transform $x = 24$ in order to use Table A.4. In this case, we would use $y = 24/4 = 6$. Thus the probability is 0.715.

- What is $P(20 < X < 40)$?

  Similarly we must adjust the range in order to use Table A.4. Instead of (20,40), we need to use (5,10). Thus, the probability is 0.971-.56=.411.

Exponential Distribution

- Special case of Gamma distribution with

$$\left.\begin{array}{l}\alpha = 1 \\ \beta = 1/\lambda\end{array}\right\} \rightarrow \begin{array}{l}\mu = 1/\lambda \\ \sigma^2 = 1/\lambda^2\end{array}$$

- Widely used in engineering disciplines

$$f(x) = \lambda \exp(-\lambda x) \text{ for } x \geq 0$$

- Can easily be integrated so

$$F(x) = 1 - \exp(-\lambda x) \text{ for } x \geq 0$$

Exponential Distribution

- Memoryless property

$$\begin{aligned} P(X \geq t + t_0 | X \geq t_0) &= \frac{P(X \geq t + t_0)}{P(X \geq t_0)} \\ &= \frac{\exp(-\lambda(t + t_0))}{\exp(-\lambda t_0)} \\ &= \exp(-\lambda t) \\ &= P(X \geq t) \end{aligned}$$

  The distribution of additional lifetime is the same as the original lifetime distribution

- Poisson process : Suppose the number of events occurring in any interval t is Poisson(αt). The time between successive events is exponential with parameter $\lambda = \alpha$

Example

An interrupt service unit takes t_0 seconds to service an interrupt before it can handle a new one. Suppose that the interrupt arrivals follow a Poisson distribution with an average of λ interrupts per second. What is the probability that an interrupt is lost?

ANSWER: An interrupt is lost if the inter-arrival time between consecutive interrupts is less than t_0 seconds. Since the time between interrupts is Exponential(λ), the probability would be

$$P(X < t_0) = 1 - \exp(-\lambda t_0)$$

Chi-Squared Distribution

- Special case of Gamma distribution with

$$\left.\begin{array}{l}\alpha = \nu/2 \\ \beta = 2\end{array}\right\} \rightarrow \begin{array}{l}\mu = \nu \\ \sigma^2 = 2\nu\end{array}$$

- Related to the normal distribution

  $Z^2 = \chi_1^2$

  Also see Problem 4.65

- Used in variety of statistical tests

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} \exp(-x/2) \text{ for } x \geq 0$$

Weibull Distribution

- Frequently used as a lifetime distribution

$$f(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \exp(-(x/\beta)^\alpha \text{ for } x \geq 0$$

- Exponential when $\alpha = 1$ and $\beta = 1/\lambda$

$$\begin{aligned}\mu &= \beta\Gamma\left(1 + \frac{1}{\alpha}\right) \\ \sigma^2 &= \beta^2\left\{\Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right)\right]^2\right\}\end{aligned}$$

- CDF has closed form

$$F(x) = 1 - \exp(-(x/\beta)^\alpha) \text{ for } x \geq 0$$

Log-Normal Distribution

- If X lognormal $\rightarrow$ log(X) is normal

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\log(x) - \mu)^2}{2\sigma^2}\right) \text{ for } x \geq 0$$

- μ and σ are not mean and std dev

$$\begin{aligned}E(X) &= \exp(\mu + \sigma^2/2) \\ V(X) &= \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)\end{aligned}$$

- Since log(X) is normal,

$$\begin{aligned}P(X \leq x) &= P(\log(X) \leq \log(x)) \\ &= P\left(Z \leq \frac{\log(x) - \mu}{\sigma}\right)\end{aligned}$$

Beta Distribution

- Unlike previous distributions, provides distribution on interval of finite length, $B - A$

$$f(x) = \frac{1}{B-A}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{x-A}{B-A}\right)^{\alpha-1}\left(\frac{B-x}{B-A}\right)^{\beta-1}$$

  for $A \leq x \leq B$

$$\begin{aligned}E(X) &= A + (B-A)\frac{\alpha}{\alpha+\beta} \\ V(X) &= \frac{(B-A)^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\end{aligned}$$

- Standard Beta when interval over [0,1]

- Commonly used to model variation in percentage or proportion of a quantity

Probability Plots

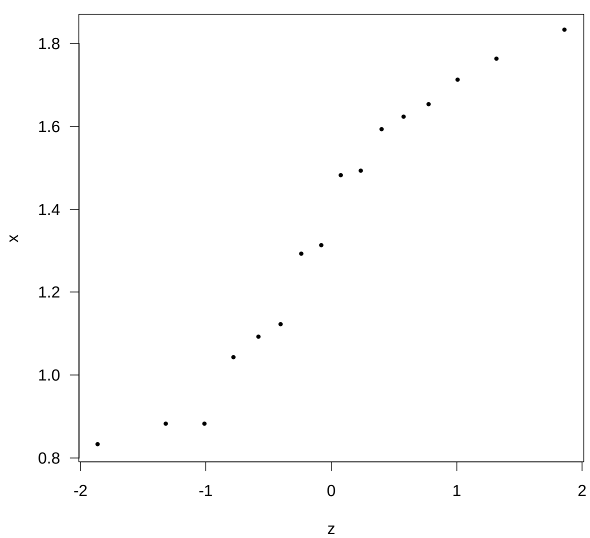
- Used to check distributional assumption
- Compare sample percentiles with distribution percentiles
- If reasonable distribution, points should follow straight line
- Sample percentile
  - Order obs from smallest to largest
  - ith smallest obs is the $100[(i - .5)/n]$%-tile
- Distribution percentile
  - Use cdf to compute appropriate %-tile
- If normal dist, known as normal prob plot

Example: Problem 83

Sample of coating thicknesses for low viscosity paint

Thickness	Percentile	z
0.83	.0312	-1.86
0.88	.0938	-1.32
0.88	.1563	-1.01
1.04	.2188	-0.78
1.09	.2813	-0.58
1.12	.3438	-0.40
1.29	.4063	-0.24
1.31	.4688	-0.08
1.48	.5313	0.08
1.49	.5938	0.24
1.59	.6563	0.40
1.62	.7188	0.58
1.65	.7813	0.78
1.71	.8438	1.01
1.76	.9063	1.32
1.83	.9688	1.86

Example: Problem 83



Probability Plots

- Similar approach can be taken with other distributions with only scale and location parameters.
- Compute percentiles from *standard distribution*
- Can do this with Weibull using $\log(X)$ vs the extreme value distribution
- For Gamma must estimate α and β before computing the percentiles
- Most software packages do this