System Simulation
Part II: Mathematical and Statistical Models
Chapter 5: Statistical Models

Fatih Cavdur
fatihcavdur@uludag.edu.tr

March 29, 2014

Introduction

- The world of the model-builder sees probabilistic rather than deterministic.
- There are many causes of variation. The time it takes to fix a broken machine is a function of the complexity of the breakdown, whether the repairman brought the proper replacement parts and tools, whether the operator receives a lesson in preventive maintenance etc.
- To the model-builder, these occur by chance and cannot be predicted, however, some statistical model might well describe the time to make a repair.

Discrete Random Variables

- If the number of possible values of X is finite or countably infinite, X is called a discrete RV.
- The possible values of X are given by the range space of X, denoted by R_X. Here, $R_X = 0, 1, \ldots$.
- Let X be a discrete RV. With each possible outcome x_i in R_X, a number, $p(x_i) = P(X = x_i)$ gives the probability that the RV equals the value of x_i.
- We have
  - $p(x_i) \geq 0$ for all i
  - $\sum_{i=1}^{\infty} p(x_i) = 1$
- $p(x_i)$ is called the probability mass function (PMF) of X.

Discrete Random Variables-cont.

- Consider the experiment of tossing a single biased die. Define X as the number of spots on the up face, $R_X = 1, 2, 3, 4, 5, 6$. Here, the probability that a given face lands up is proportional to the number of spots showing. So, the discrete probability distribution of of this experiment is given by

x_i	1	2	3	4	5	6
$p(x_i)$	$1/21$	$2/21$	$3/21$	$4/21$	$5/21$	$6/21$

Discrete Random Variables-cont.

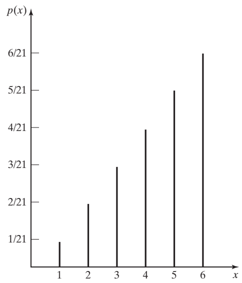

Figure : PMF for the Die Example

Continuous Random Variables

- Continuous Random Variables (Continuous RV): If the range space R_X of the RV X is an interval or a collection of intervals, X is called a continuous RV.
- We have

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

- We also have
  - $f(x) \geq 0$ for all x in R_X
  - $\int_{R_X} f(x)dx = 1$
  - $f(x) = 0$ if x is not in R_X

Continuous Random Variables-cont.

- $f(x)$ is called the probability density function (PDF) of the RV X. Also,

$$P(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$$

- and,

$$\begin{aligned} P(a \le X \le b) &= P(a < X \le b) \\ &= P(a \le X < b) \\ &= P(a < X < b) \end{aligned}$$

Continuous Random Variables-cont.

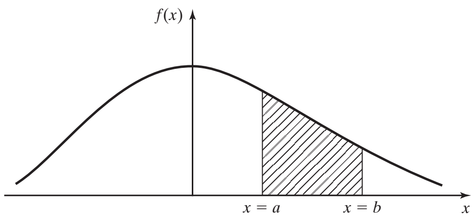

Figure : Graphical Interpretation of $P(a < X < b)$

Continuous Random Variables-cont.

- The life of a device used to inspect cracks in aircraft wings is given by X, a continuous RV with a PDF of

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

- The probability that the life of the device is between 2 and 3 years is

$$P(2 \leq X \leq 3) = \frac{1}{2}\int_2^3 e^{-x/2}dx = -e^{-3/2} + e^{-1} = 0.145$$

Continuous Random Variables-cont.

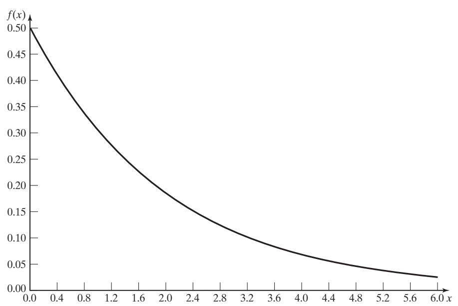

Figure : PDF for the Example Problem

Cumulative Distribution Function

- The cumulative distribution function (CDF), denoted by $F(x)$, shows the probability that $F(x) = P(X \leq x)$.
- If X is discrete,

$$f(x) = \sum_{x_i \leq x} p(x_i)$$

- If X is continuous,

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Cumulative Distribution Function-cont.

- Some properties of the CDF are
  - F is a non-decreasing function.
  - $\lim_{x \to -\infty} F(x) = 0$
  - $\lim_{x \to \infty} F(x) = 1$
- All probability questions about X can be answered in terms of the CDF, such as

$$P(a < X \leq b) = F(b) - F(a), \text{ for all } a < b$$

- Look at the examples in your text.

Expectation and Variance

- Expectation of X if X is discrete,

$$E(X) = \sum_i x_i p(x_i)$$

- Expectation of X if X is continuous,

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

- Variance of X ($\sigma^2, \mathrm{Var}(X), V(X)$)

$$V(X) = E\left([X - E(X)]^2\right) = E\left(X\right)^2 - [E(X)]^2$$

Probability Distributions

- Consider the single-server-queuing system. The arrivals and service times might be deterministic or probabilistic.
- We can use probability distributions to model the stochastic features. For example, for service times
  - if the data are completely random, the exponential distribution might be considered
  - if the data fluctuate from some value, the normal or the truncated normal distribution might be used
  - the gamma and Weibull distributions can also be used to model inter-arrival and service times
- Some others, such as the geometric, Poisson and negative binomial distributions might be considered for demand distribution etc.

Bernoulli Distribution

- Consider an experiment with n trials each of which can be a success of failure with the corresponding probabilities. It is called a Bernoulli process if the trials are independent and the probability of success is the same for all trials. We have

$$p_j(x_j) = p(x_j) = \begin{cases} p, & x_j = 1,\ j = 1, 2, \dots, n \\ 1 - p = q, & x_j = 0,\ j = 1, 2, \dots, n \\ 0, & \text{otherwise} \end{cases}$$

$$E(X_j) = 0 \times q + 1 \times p = p$$

$$V(X_j) = 0^2 \times q + 1^2 \times p - p^2 = pq$$

Binomial Distribution

- PDF

$$p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x}, & x = 0, 1, 2, \dots, n \\ 0, & \text{otherwise} \end{cases}$$

- Expectation

$$E(X) = np$$

- Variance

$$V(X) = npq$$

Binomial Distribution-Example

- A production process manufactures computer chips that are defective 2% of the time on average. Every day a random sample of size 50 is taken from the process. If the sample contains more than 2 defectives, the process is stopped. Compute the probability of the process stopping.

Binomial Distribution-Example

- If we let X be the number of defectives in the sample, then, X have a binomial distribution with parameters $(n = 50, p = 0.02)$. Hence,

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - \sum_{x=0}^{2} \binom{50}{x} (0.02)^x (0.98)^{50-x} \\ &= 0.92 \end{aligned}$$

Binomial Distribution-Example

- The mean number of defectives is

$$E(X) = np = 50 \times 0.02 = 1$$

- Variance

$$V(X) = npq = 50 \times 0.02 \times 0.98 = 0.98$$

Geometric Distribution

- PDF

$$p(x) = \begin{cases} pq^{x-1}, & x = 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

- Expectation

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

- Variance

$$V(X) = \frac{q}{p^2}$$

Negative Binomial Distribution

- PDF

$$p(x) = \begin{cases} \binom{y-1}{k-1} p^k q^{y-k}, & y = k, k+1, k+2, \dots \\ 0, & \text{otherwise} \end{cases}$$

- Expectation

$$E(X) = \frac{k}{p}$$

- Variance

$$V(X) = \frac{kq}{p^2}$$

Geometric-Negative Binomial Distribution-Example

- 40% of the assembled ink-jet printers are rejected at the inspection station. Find the probability that the first acceptable ink-jet printer is the third one inspected.
- If we let X be the number of trials to achieve the first success (acceptable printer),

$$p(3) = (0.4)^2(0.6) = 0.096$$

- The probability that the third printer inspected is the second acceptable printer is

$$p(3) = \binom{3-1}{2-1}(0.4)^{3-2}(0.6)^2 = 0.288$$

Poisson Distribution

- PDF and CDF

$$p(x) = \begin{cases} \frac{e^{-\alpha}\alpha^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

$$F(X) = \sum_{i=0}^{x} \frac{e^{-\alpha}\alpha^i}{i!}$$

- Expectation and Variance

$$E(X) = \alpha$$

$$V(X) = \alpha$$

Poisson Distribution-cont.

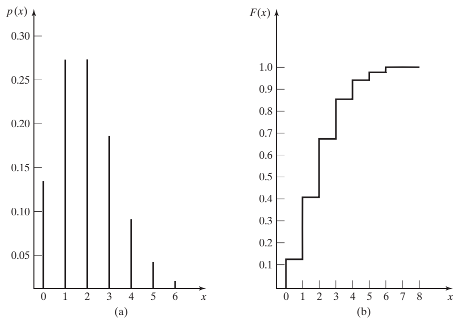

Figure : Poisson PMF and CDF

Poisson Distribution-Example

- A computer repairman is "beeped" each time there is a call for service. The number of beeps per hour is a Poisson RV with a mean of $\alpha = 2$ per hour. The probability of 3 beeps in the next hour is given by

$$p(3) = \frac{e^{-2}(2)^3}{3!} = 0.18 = F(3) - F(2) = 0.857 - 0.677$$

- The probability of 2 or more beeps in a 1-hour period is

$$P(X \geq 2) = 1 - P(X < 2) = 1 - F(1) = 0.594$$

Uniform Distribution

- PDF and CDF

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

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

- Expectation and Variance

$$E(X) = \frac{a+b}{2}$$

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

Uniform Distribution-cont.

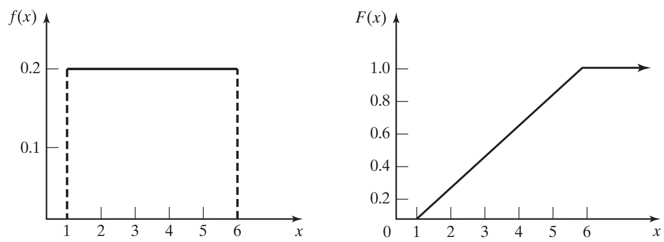

Figure : Uniform PDF and CDF

Exponential Distribution

- PDF and CDF

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

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

- Expectation and Variance

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

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

Exponential Distribution-cont.

- One of the most important properties of the exponential distribution is that it is "memoryless", which means, for all $s \geq 0$ and $t \geq 0$,

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

Exponential Distribution-cont.

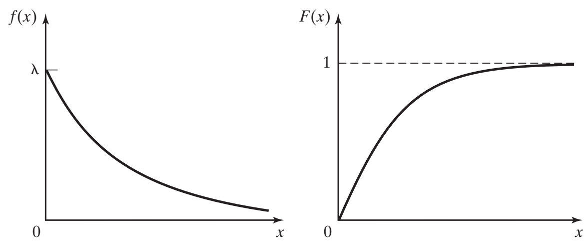

Figure : Uniform PDF and CDF

Exponential Distribution-cont.

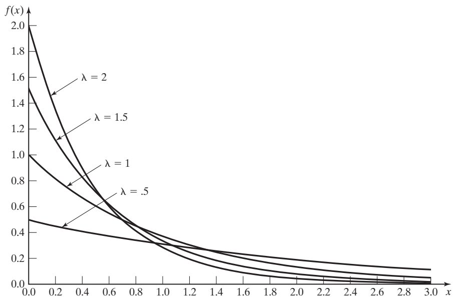

Figure : PDFs for Several Exponentials

Exponential Distribution-Example

- The life time of a lamp, in thousand hours, is exponentially distributed with failure rate $\lambda = 1/3$. Find the probability that the lamp will last for another 1,000 hours given that it is working after 2,500 hours.

$$
\begin{aligned}
P(X > 3.5 | X > 2.5) &= P(X > 1) \\
&= e^{-1/3} \\
&= 0.717
\end{aligned}
$$

Gamma Distribution

- PDF and CDF

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

$$F(X) = \begin{cases} 1 - \int_x^\infty \frac{\beta\theta}{\Gamma(\beta)}(\beta\theta t)^{\beta-1}e^{-\beta\theta t}dt, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

- Expectation and Variance

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

$$V(X) = \frac{1}{\beta\theta^2}$$

Gamma Distribution-cont.

- Gamma Function

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

- We can show that

$$\Gamma(\beta) = (\beta - 1)!$$

- When $\beta = k$, the distribution is called as the Erlang distribution of order k.

Gamma Distribution-cont.

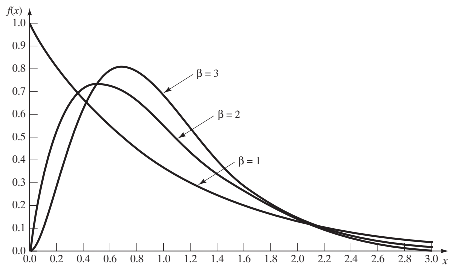

Figure : PDFs for Several Gammas

Normal Distribution

- PDF and CDF

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

$$F(X) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] dt, \quad -\infty < x < \infty$$

- It is difficult to compute the above integral analytically. We can do it numerically, but it would be dependent on the mean and the variance. Not to be so, we can transform it to standard normal distribution.

Normal Distribution-cont.

- Let $z = (t - \mu)/\sigma$. Then,

$$\begin{aligned} F(X) = P(X \leq x) &= P\left(Z \leq \frac{x-\mu}{\sigma}\right) \\ &= \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\ &= \int_{-\infty}^{(x-\mu)/\sigma} \phi(z) dz \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

Normal Distribution-cont.

- Hence, $X\ N(\mu, \sigma^2)$ and $Z\ N(0, 1)$. The PDF of Z is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \quad -\infty < x < \infty$$

- The CDF is computed and tabulated for use.

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

Normal Distribution-cont.

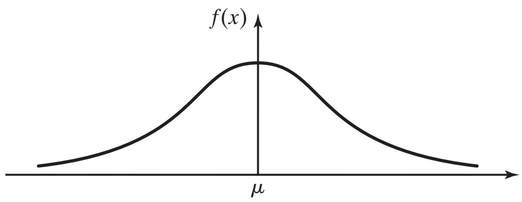

Figure : PDF for Normal

Normal Distribution-cont.

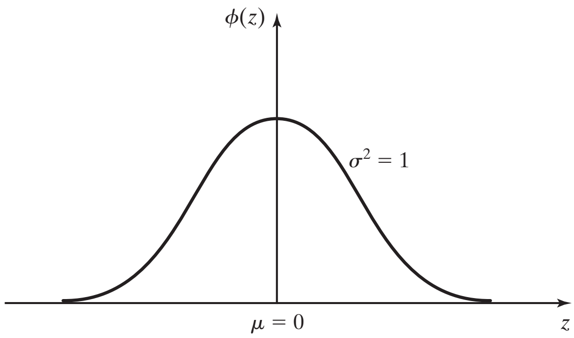

Figure : PDF for Standard Normal

Normal Distribution-cont.

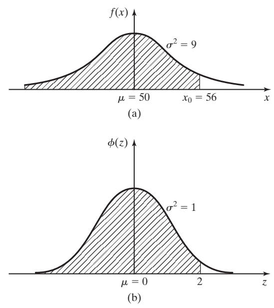

Figure : from Normal to Standard Normal

Normal Distribution-Example

- X is approximated by a normal distribution with mean 25 and variance 9. Compute the value for X that will be exceeded only 5% of the time.

$$P(X > x_0) = P\left(Z > \frac{x_0 - 25}{3}\right) = 1 - \Phi\left(\frac{x_0 - 25}{3}\right)$$

- Hence,

$$1 - \Phi\left(\frac{x_0 - 25}{3}\right) = 0.05 \Rightarrow \Phi\left(\frac{x_0 - 25}{3}\right) = 0.95$$

$$\Phi(1.645) = 0.95 \Rightarrow \frac{x_0 - 25}{3} = 1.645 \Rightarrow x_0 = 29.935$$

Normal Distribution-Example

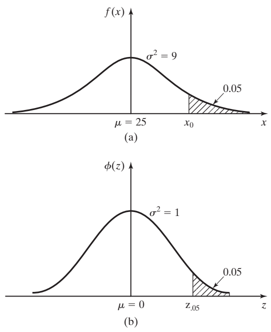

Figure : Normal Distribution Example

Weibull Distribution-cont.

- PDF and CDF

$$f(x) = \begin{cases} \frac{\beta}{\alpha}\left(\frac{x-v}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x-v}{\alpha}\right)^{\beta}\right], & x \geq v \\ 0, & \text{otherwise} \end{cases}$$

$$F(X) = \begin{cases} 1 - \exp\left[-\left(\frac{x-v}{\alpha}\right)^{\beta}\right], & x \geq v \\ 0, & \text{otherwise} \end{cases}$$

Weibull Distribution-cont.

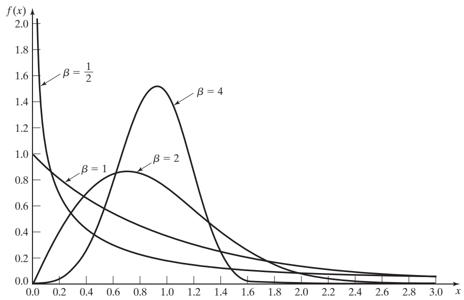

Figure : Several PDFs for Weibull

Weibull Distribution-cont.

- Expectation and Variance

$$E(X) = v + \alpha\Gamma\left(\frac{1}{\beta} + 1\right)$$

$$V(X) = \alpha^2 \left(\Gamma\left(\frac{2}{\beta} + 1\right) - \left[\Gamma\left(\frac{1}{\beta} + 1\right)\right]^2\right)$$

Triangular Distribution

- PDF and CDF

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)}, & a \le x \le b \\ \frac{2(c-x)}{(c-b)(c-a)}, & b < x \le c \\ 0, & \text{otherwise} \end{cases}$$

$$F(X) = \begin{cases} 0, & x \le a \\ \frac{(x-a)^2}{(b-a)(c-a)}, & a < x \le b \\ 1 - \frac{(c-x)^2}{(c-b)(c-a)}, & b < x \le c \\ 1, & x > c \end{cases}$$

System Simulation Part II: Mathematical and Statistical Models Chapter 5: Statistical Models
└Continuous Distributions

Triangular Distribution-cont.

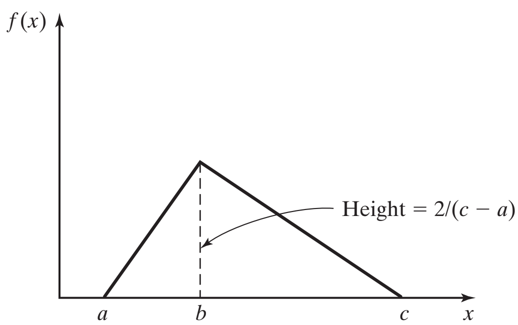

Figure : PDF for Triangular

Triangular Distribution-cont.

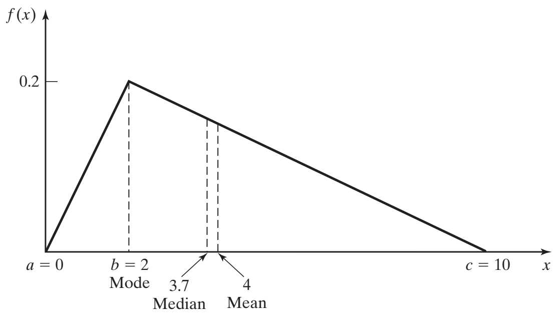

Figure : Mode, Median and Mean

Lognormal Distribution

- PDF

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right], & x \geq 0 \\ \frac{2(c-x)}{(c-b)(c-a)}, & b < x \leq c \\ 0, & \text{otherwise} \end{cases}$$

- Expectation and Variance

$$E(X) = e^{\mu+\sigma^2/2}$$

$$V(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$$

Lognormal Distribution-cont.

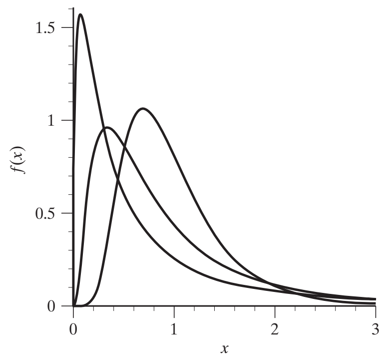

Figure : PDF for Lognormal

Beta Distribution

- PDF

$$f(x) = \begin{cases} \frac{x^{\beta_1 - 1}(1-x)^{\beta_2 - 1}}{B(\beta_1, \beta_2)}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

where

$$B(\beta_1, \beta_2) = \Gamma(\beta_1)\Gamma(\beta_2)/\Gamma(\beta_1 + \beta_2)$$

Beta Distribution-cont.

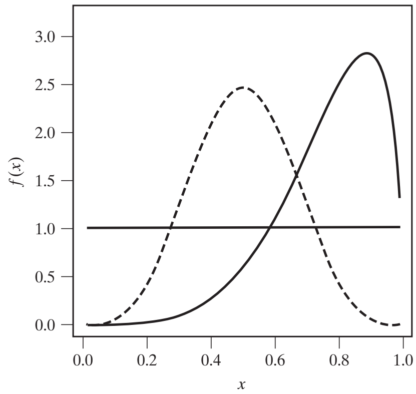

Figure : Several PDFs for Beta

Poisson Process

The counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process with mean rate λ if

- arrivals occur one at a time
- $\{N(t), t \geq 0\}$ has stationary increments
- $\{N(t), t \geq 0\}$ has independent increments

We can show that

$$P[N(t) = n] = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \quad \text{for } t \geq 0 \text{ and } n = 0, 1, 2, \ldots$$

which is a Poisson distribution with parameter $\alpha = \lambda t$. Hence, the mean and variance are both $\alpha = \lambda t$. We can show that the inter-arrival times are exponential.

Empirical Distributions

- An empirical distribution might be discrete or continuous.
- Its parameters are the observed values in a sample dataset.
- We might use an empirical distribution when it is not possible or not necessary to establish an RV has a particular parametric distribution.
- One advantage of an empirical distribution is that no assumption is needed beyond the observed values, but it is also a disadvantage because the sample might not cover the entire range of possible values.

Summary

- Reading HW: Chapter 5.
- Chapter 5 Exercises.