Important Probability Distributions

Professor Yongmiao Hong
Cornell University
June 25, 2019

CONTENTS

Introduction

- A probability space $(S, \mathbb{B}, P)$ completely characterizes a random experiment.
- It is usually assumed that there exists a unique true but unknown probability distribution for the random experiment of interest.
- One often considers a class of probability distributions, say PMF/PDF $f(x, \theta)$, indexed by parameter θ, where the functional form $f(\cdot, \cdot)$ is known.

Introduction

- Different values of θ yield different distributions. The collection of these distributions constitute a family of probability distributions.
- This family of probability distributions is called a class of parametric probability distribution models.
- One main objective of statistics and econometrics is to use observed economic data to estimate the unknown true parameter value, say θ_0, under the assumption that there exists some parameter value θ_0 such that $f(x, \theta_0)$ coincides with the true probability distribution $f_X(x)$ of the random experiment, namely $f(x, \theta_0) = f_X(x)$.

Introduction

- This chapter introduces a variety of important distribution models, and discuss their properties and applications in economics and finance.
- We emphasize that it is important to understand the meanings and roles that parameters play in each parametric distribution.

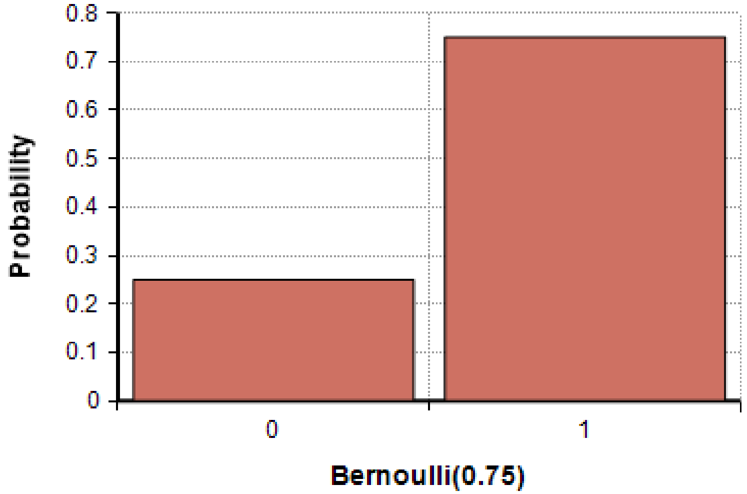
CONTENTS

Bernoulli Distribution

A DRV X follows a Bernoulli(p) distribution if its PMF

$$\begin{aligned} f_X(x) &= \begin{cases} p & \text{if } x = 1, \\ 1-p & \text{if } x = 0, \end{cases} \\ &= p^x(1-p)^{1-x} \text{ for } x = 0, 1, \end{aligned}$$

where $0 < p < 1$.



Remark:

This is a binary random variable, taking value 1 with probability p, and taking value 0 with probability $1-p$.

Bernoulli Distribution

Remarks:

- For Bernoulli(p),

$$\begin{aligned} E(X) &= p, \\ \text{var}(X) &= p(1-p). \end{aligned}$$

- MGF of Bernoulli(p)

$$M_X(t) = pe^t + 1 - p \text{ for } -\infty < t < \infty.$$

- Because $p = P(X = 1)$ and X only takes two possible values, parameter p fully characterizes the probability distribution of Bernoulli(p).

Bernoulli Distribution

Remarks:

- Bernoulli(p) has wide applications in economics and finance:
  - Neftci (1984) examines the asymmetry of U.S. business cycles by considering the time series properties of signed processes from a positive change to a negative change, and from a negative change to a positive change.

Bernoulli Distribution

Remarks:

- Das (2002) approximates jumps of interest rates by a sequence of Bernoulli random variables.
- Logit regression model

$$P(X = 1|Z) = \frac{1}{1 + \exp(-\beta' Z)}.$$

Important applications of logit regression include pattern recognition and classification analysis.

Binomial Distribution

A DRV X follows a binomial distribution, denoted as $B(n, p)$, if its PMF

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots, n.$$

where $n \geq 1$ and $0 < p < 1$.

Binomial Distribution

Remarks:

- A binomial random variable can take $n+1$ possible integer values $\{0, 1, \cdots, n\}$.

- Question: When can this distribution arise?

  Suppose one throws a coin n times independently. How many heads can one get from these n trials?

Binomial Distribution

Remarks:

- Let X_i = the number of heads for the i-th trial, and let X = the total number of heads in the n trials. Then

$$X = \sum_{i=1}^{n} X_i,$$

where X_i is a sequence of IID Bernoulli(p) random variables. It can be shown that X follows a $B(n, p)$ distribution.

To be Continued

Binomial Distribution

Question: How to verify

$$\sum_{x=0}^{n} f_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = 1 \text{ for all } n \geq 1 \text{ and all } p \in (0,1)?$$

- By the binomial theorem that for any real x and y and integer $n \geq 0$,

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i},$$

we can obtain the identity by setting $x = p$ and $y = 1-p$. This is the reason why the distribution is called the binomial distribution.

To be Continued

Binomial Distribution

- Mean of $B(n,p)$:

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{n} x f_x(x) \\
&= \sum_{x=0}^{n} x\binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x\binom{n}{x} p^x (1-p)^{n-x} + 0 \cdot \binom{n}{0} p^0 (1-p)^{n-0} \\
&= \sum_{x=1}^{n} n\binom{n-1}{x-1} p^x (1-p)^{n-x} \\
&= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{(n-1)-y} \\
&= np \sum_{y=0}^{n-1} f_Y(y) = np.
\end{aligned}
$$

To be Continued

Binomial Distribution

- Variance of $B(n,p)$:

$$
\begin{aligned}
E(X^2) &= \sum_{x=0}^{n} x^2 f_X(x) \\
&= \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \cdot x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} xn \binom{n-1}{x-1} p^x (1-p)^{n-x} \\
&= n \sum_{y=0}^{n-1} (y+1) \binom{n-1}{y} p^{y+1} (1-p)^{(n-1)-y}
\end{aligned}
$$

$$
\begin{aligned}
&= np \sum_{y=0}^{n-1} y \binom{n-1}{y} p^y (1-p)^{(n-1)-y} \\
&\quad + np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{(n-1)-y} \\
&= npE(Y) + np \sum_{y=0}^{n-1} f_Y(y) \\
&= np(n-1)p + np \\
&= np[(n-1)p + 1].
\end{aligned}
$$

To be Continued

Binomial Distribution

- Variance of $B(n,p)$:

It follows that

$$\begin{aligned} \sigma_X^2 &= E(X^2) - \mu_X^2 \\ &= \{np \cdot [(n-1)p] + np\} - (np)^2 \\ &= np(1-p). \end{aligned}$$

To be Continued

Binomial Distribution

- MGF of $B(n,p)$:

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\
&= (pe^t + 1 - p)^n.
\end{aligned}
$$

Binomial Distribution

- $B(n, p)$ has been one of the oldest to have been the subject of study in probability and statistics.

  It has wide applications in economics:

  - In quality control, for example, it can be used to approximate the distribution of the numbers of defective products in a total of n products each of which is probability p to be defective.
  - It can also be used to model the cumulative number of jumps occurred in financial price movements during a given period of time (e.g., Das 2002).

Negative Binomial Distribution

- $B(n, p)$ describes the probability distribution for the number of successes in a fixed number of n trials.
- Negative Binomial, $NB(n, p)$, characterizes the probability distribution of the number of trials required to obtain a given number of successes.
- In a sequence of independent Bernoulli(p) trials, let $X =$ the number of trials such that at the X-th trial the r-th success occurs, where r is a fixed integer. This implies that $X - 1$ is the number of trials right before the r-th success is obtained.

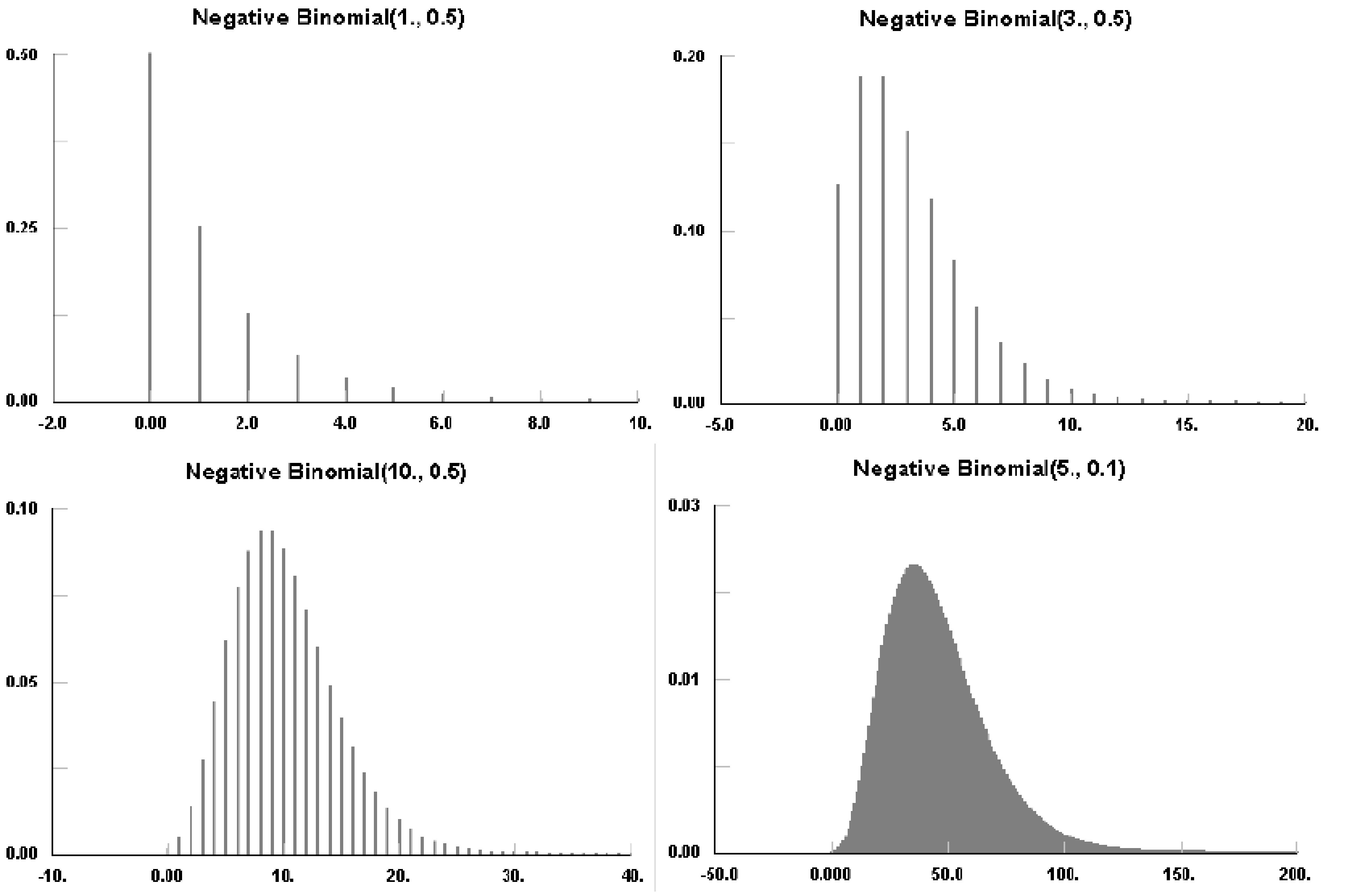
To be Continued

Negative Binomial Distribution

- There are $r-1$ successes in the first $X-1$ trials and the X-th trial is a success, so the PMF of X

$$
\begin{aligned}
f_X(x) &= \left[\binom{x-1}{r-1} p^{r-1}(1-p)^{(x-1)-(r-1)}\right] \cdot p \\
&= \binom{x-1}{r-1} p^r (1-p)^{x-r}, \qquad x = r, r+1, \cdots
\end{aligned}
$$

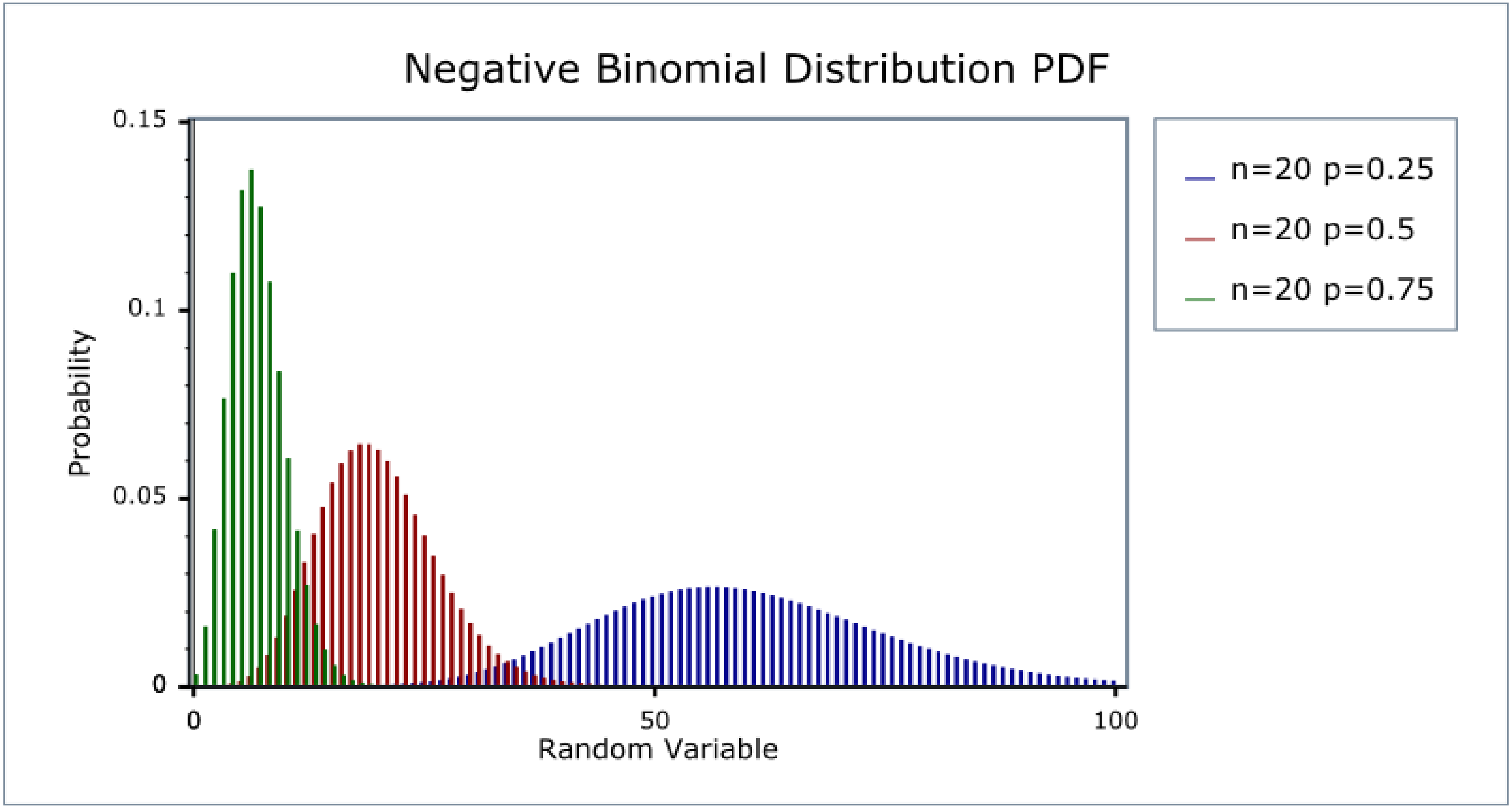
To be Continued


Negative Binomial(1., 0.5)
0.50
0.25
0.00
-2.0
0.00
2.0
4.0
6.0
8.0
10.
Negative Binomial(3., 0.5)
0.20
0.10
0.00
-5.0
0.00
5.0
10.
15.
20.
Negative Binomial(10., 0.5)
0.10
0.05
0.00
-10.
0.00
10.
20.
30.
40.
Negative Binomial(5., 0.1)
0.03
0.01
0.00
-50.0
0.000
50.0
100.
150.
200.

Negative Binomial Distribution

- $NB(n,p)$ is sometimes defined in terms of the number of failures when obtaining the r-th success, $Y = X - r$.
  - The support of Y is the set of all nonnegative integers $\{0, 1, \cdots\}$.
  - The PMF of Y

$$
\begin{aligned}
f_Y(y) &= P(Y = y) \\
&= P(X = y + r) \\
&= \binom{y+r-1}{r-1} p^r (1-p)^y, \text{ for } y = 0, 1, \cdots .
\end{aligned}
$$


Negative Binomial Distribution PDF
Probability
0.15
0.1
0.05
0
0
50
100
Random Variable
n=20 p=0.25
n=20 p=0.5
n=20 p=0.75

Geometric Distribution

The geometric distribution is the probability distribution of the number of Bernoulli trials required to obtain the first success.

This is a special case of the negative binomial distribution with $r = 1$.

When $r = 1$, the negative binomial distribution becomes

$$f_X(x) = p(1-p)^{x-1}, \qquad x = 1, 2, \cdots.$$

Geometric Distribution

Remarks:

- The geometric distribution is the simplest of the waiting time distributions. The random variable X is the number of trials required to obtain the first success, so we are "waiting for a success".

Geometric Distribution

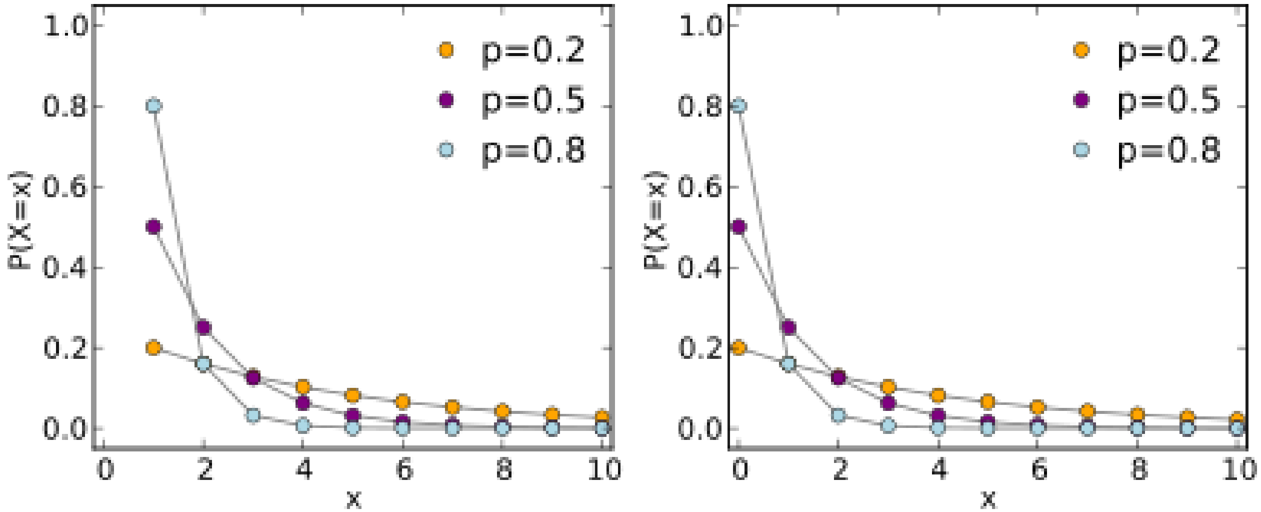
Remarks:

- The geometric distribution has the so-called "memoryless" or "nonaging" property: for integers $s > t$, we have

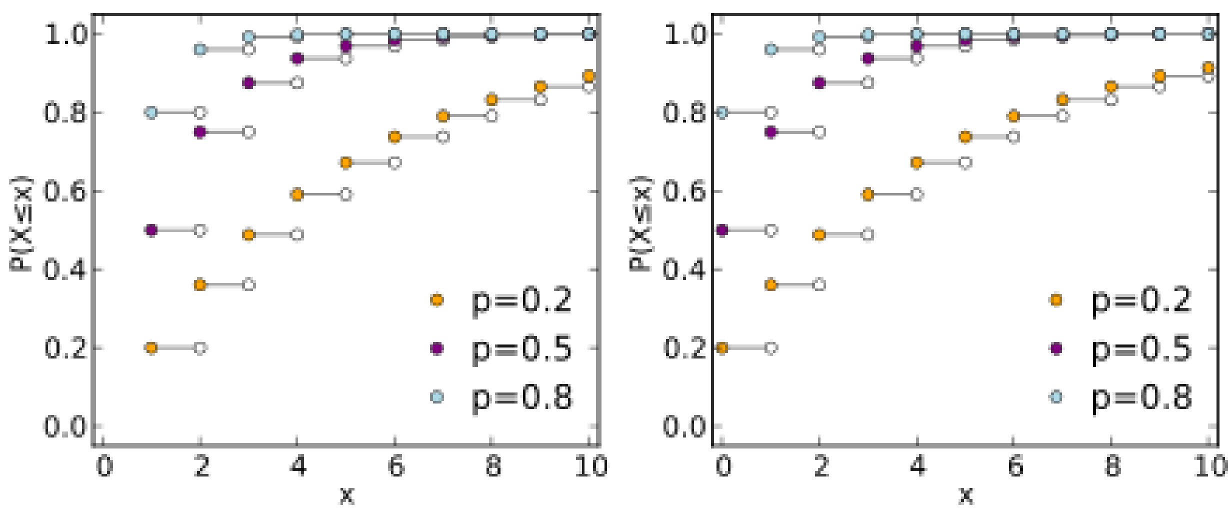
$$P(X > s | X > t) = P(X > s - t).$$

  That is, the probability of getting additional $s-t$ failures, having already observed t failures, is the same as the probability of observing $s - t$ failures at the start of the sequence. In other words. the probability of getting a run of failures depends only on the length of the run, not on its position.

Probability mass function



Cumulative distribution function



Geometric Distribution

Question: How to show the "memoryless" property?

- Using the formula $P(A|B) = P(A \cap B)/P(B)$ and the fact that when $s > t$, the event $\{X > t\}$ contains the event $\{X > s\}$ as a subset, we have

To be Continued

Geometric Distribution

$$
\begin{aligned}
P(X > s | X > t) &= \frac{P(X > s, X > t)}{P(X > t)} \\
&= \frac{P(X > s)}{P(X > t)} \\
&= \frac{1 - P(X \leq s)}{1 - P(X \leq t)} \\
&= \frac{1 - \sum_{x=1}^{s} p(1-p)^{x-1}}{1 - \sum_{x=1}^{t} p(1-p)^{x-1}} \\
&= \frac{(1-p)^s}{(1-p)^t} \\
&= (1-p)^{s-t} \\
&= P(X > s - t).
\end{aligned}
$$

To be Continued

Geometric Distribution

- The geometric distribution is usually viewed as a discrete analog of the exponential distribution to be introduced later.
- It can be used to model births and populations.

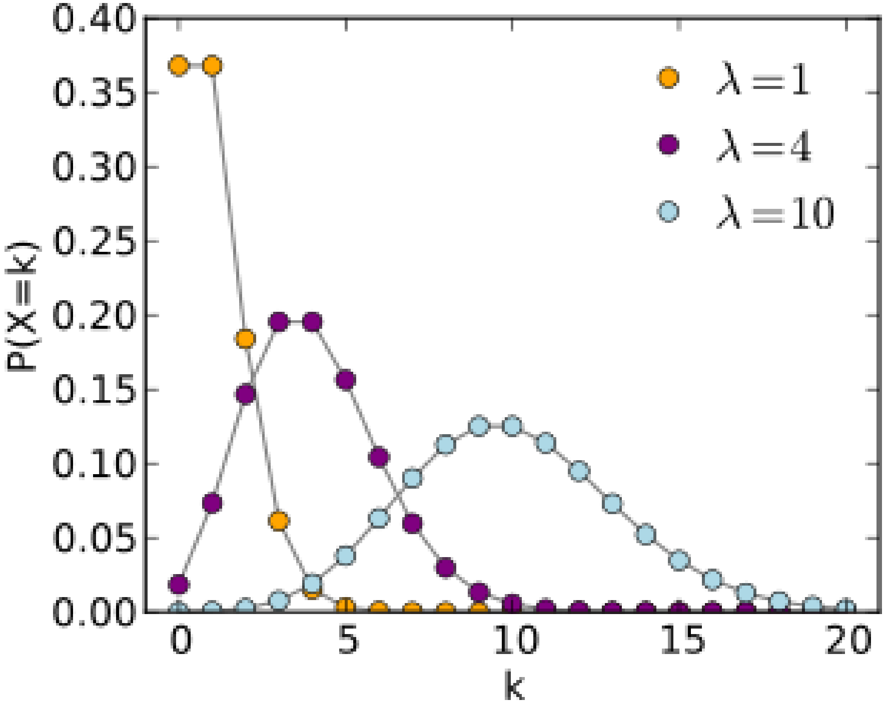
Poisson Distribution

A DRV X follows a Poisson(λ) distribution if its PMF

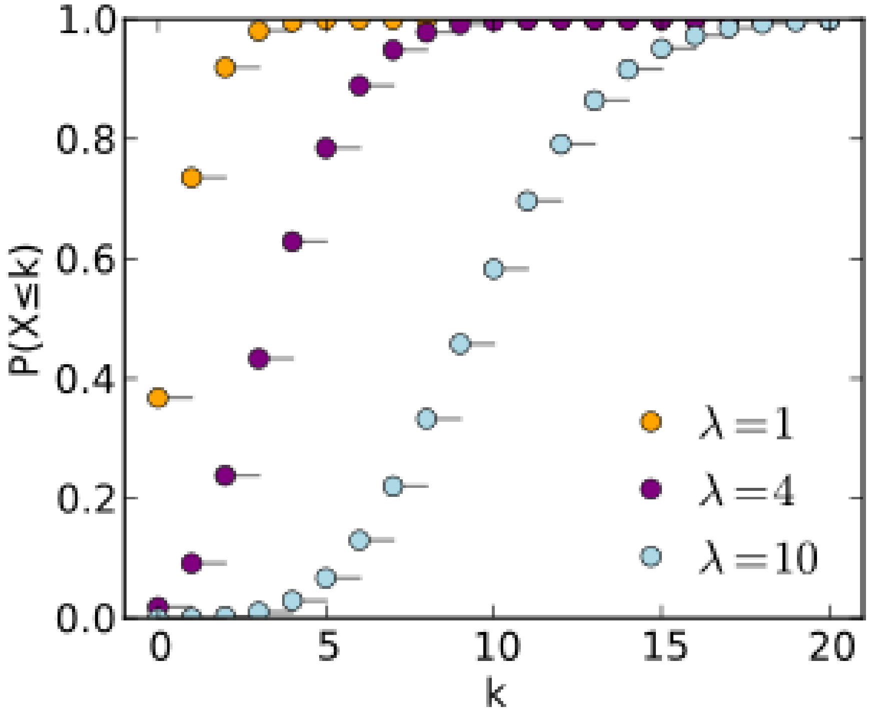
$$f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, \cdots,$$

where $\lambda > 0$.
The parameter λ is called an intensity parameter.

Probability mass function



Cumulative distribution function



Poisson Distribution

Remarks:

- The support of Poisson(λ) is the set of all nonnegative integers.
- Using MacLaurin's series expansion

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!},$$

we have

$$\begin{aligned} \sum_{x=0}^{\infty} f_X(x) &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} e^{\lambda} \\ &= 1\,. \end{aligned}$$

To be Continued

Poisson Distribution

- Mean of Poisson(λ):

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} \\ &= \lambda \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda \sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} \\ &= \lambda. \end{aligned}$$

- Interpretation for $\lambda = E(X)$: λ is the average number for the occurring events following Poisson(λ) in a unit of time period.

To be Continued

Poisson Distribution

- Variance of Poisson(λ):

$$
\begin{aligned}
E(X^2) &= \sum_{x=0}^{\infty} x^2 e^{-\lambda} \frac{\lambda^x}{x!} \\
&= \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^x}{(x-1)!} \\
&= \lambda \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda \sum_{y=0}^{\infty} (y+1) e^{-\lambda} \frac{\lambda^y}{y!} \\
&= \lambda \sum_{y=0}^{\infty} y e^{-\lambda} \frac{\lambda^y}{y!} + \lambda \sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} \\
&= \lambda E(Y) + \lambda \sum_{y=0}^{\infty} f_Y(y) \\
&= \lambda^2 + \lambda,
\end{aligned}
$$

it follows that

$$
\begin{aligned}
\sigma_X^2 &= E(X^2) - \mu_X^2 \\
&= (\lambda^2 + \lambda) - \lambda^2 \\
&= \lambda.
\end{aligned}
$$

To be Continued

Poisson Distribution

- Both mean and variance are equal to λ for Poisson(λ).
- MGF of Poisson(λ) :

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \sum_{x=0}^{\infty} e^{tx} f_X(x) \\
&= \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \\
&= e^{-\lambda} e^{(\lambda e^t)} \\
&= e^{\lambda(e^t - 1)}, -\infty < t < \infty.
\end{aligned}
$$

To be Continued

Poisson Distribution

- Law of Small Numbers (LSN): When $n \to \infty$ but $np \to \lambda$,

$$\begin{aligned} M_B(t) &= \left(pe^t + 1 - p\right)^n \\ &= \left[1 + \frac{np(e^t - 1)}{n}\right]^n \\ &\to e^{\lambda(e^t - 1)} = M_P(t), \end{aligned}$$

  given that $(1 + \frac{a}{n})^n \to e^a$ as $n \to \infty$.

- For large n and small p such that np is moderate, $B(n, p)$ is approximately equivalent to Poisson(λ).

- Poisson (1837) derived Poisson(λ) by considering the limit of a sequence of binomial distributions.

To be Continued

Poisson Distribution

- This is called the *law of small numbers* by the German-Russian mathematician L. V. Bortkiewicz (1898). It is also called the law of rare events or Poisson limit theorem in probability theory.

- When $r \to \infty$ but $r(1-p) \to \lambda \in (0, \infty)$, $NB(r, p)$ will also become Poisson(λ), because

$$\begin{aligned} f_Y(y) &= \binom{y+r-1}{r-1} p^r (1-p)^y \\ &\to \frac{e^{-\lambda}\lambda^y}{y!}, \qquad y = 0, 1, \cdots. \end{aligned}$$

To be Continued

Poisson Distribution

Poisson Process–A Related Concept:

- The Poisson distribution often appears in connection with the study of sequences of random events occurring over time or space.
- Suppose, starting from a time point $t = 0$, we start counting the number of random events. Then for each value of t, we obtain an integer denoted by $N(t)$, which is the number of events that have occurred during the time period $[0, t]$.

To be Continued

Poisson Distribution

- For each value of t, $N(t)$ is a discrete random variable with the set of possible values $\{0, 1, 2, ...\}$. To study the distribution of $N(t)$, the number of events occurring in $[0, t]$, we make following assumptions:
  - Stationarity: For all $m \geq 0$, and for any two equal time intervals Δ_1 and Δ_2, the probability of m events in Δ_1 is equal to the probability of m events in Δ_2;

To be Continued

Poisson Distribution

– Independent Increments: For all $m \geq 0$, and for all any time interval $(t, t+s)$, the probability of m events in $(t, t+s)$ is independent of how many events have occurred earlier or how they have occurred. In particular, suppose the times $0 \leq t_1 < t_2 < \cdots < t_k$ are given. For $1 \leq i \leq k-1$, let A_i be the event that m_i events of the process occur in $[t_i, t_{i+1})$. The independent increments mean that $\{A_1, A_2, ..., A_{k-1}\}$ is an independent set of events;

To be Continued

Poisson Distribution

– Sequencing: The occurring of two or more events in a very small time interval is practically impossible. This condition is mathematically expressed by $\lim_{\delta \to \infty} P[N(\delta) > 1]/\delta = 0$. This implies that as $\delta \to 0$, the probability of two or more events, $P[N(\delta) > 1]$, approaches 0 faster than δ does.

To be Continued

Poisson Distribution

- A stochastic process $\{N(t)\}$ that satisfies these three assumptions is called a stationary Poisson process.
  - By stationarity, the random variables $N(t_2) - N(t_1)$ and $N(t_2 + s) - N(t_1 + s)$ have the same probability distribution. The number of events in the time interval $[t_i, t_{i+1}), N(t_{i+1}) - N(t_i)$, is called the increment in the process $\{N(t)\}$ between t_i and t_{i+1}. It is worthwhile to note that stationarity and sequencing mean that the simultaneous occurrence of two or more events is impossible.

To be Continued

Poisson Distribution

- Suppose $N(0) = 0$, and we are interested in counting the number of events in the time interval $[0, t]$. We first divide the time interval $[0, t]$ into n subintervals of equal size t/n, where n is large. Then $N(t)$ is the sum of n independent Bernoulli(p) random variables with p approximately equal to $\lambda(t/n)$ for some positive constant λ. One of the basic assumptions on which the Poisson distribution is built is that, for small time intervals, the probability of an arrival is proportional to the length of waiting time.
- By the Poisson distribution approximation to the binomial distribution, we have

$$P\left[N(t) = m\right] = \frac{(\lambda t)^m e^{-\lambda t}}{m!}, \qquad m = 0, 1, 2, \ldots$$

To be Continued

Poisson Distribution

- The parameter $\lambda = E[N(1)]$ is the average number of events during a unit time period. Any process with this property is called a stationary Poisson process with rate λ, and is denoted by $\{N(t), t > 0\}$.

- It has been claimed in Douglas (1980) that the Poisson distribution plays a similar role with respect to discrete distributions to that of the normal distribution for continuous distributions.

- The Poisson distribution, together with the binomial distribution and the normal distribution, is one of the three most important probability distributions in probability and statistics.

To be Continued

Poisson Distribution

- The Poisson approximation for the binomial distribution makes the Poisson distribution has been used to model the distribution of the number of events in a specific period of time or a specific unit of space (e.g., the number of customers passing through a cashier counter, the number of telephone calls, the number of accidents on a road, the number of earthquakes in a region, the number of defaults or bankruptcies in one sector, and the number of jumps in an asset price).

To be Continued

Poisson Distribution

- Bortkiewicz (1898) used the Poisson distribution to characterize the number of deaths from kicks by horses per annum in the Prussian Army Corps where the probability of death from this cause is small while the number of soldiers exposed to the risk was large.
- A jump-diffusion model in finance: Merton (1976) proposes a model where in addition to a Brownian motion component, the price process of the underlying asset is assumed to have a Poisson jump component. This is the so-called jump diffusion model for financial derivatives pricing, which is a form of mixture model, mixing a Poisson jump process and a diffusion process.

To be Continued

Poisson Distribution

- Poisson regression in econometrics:

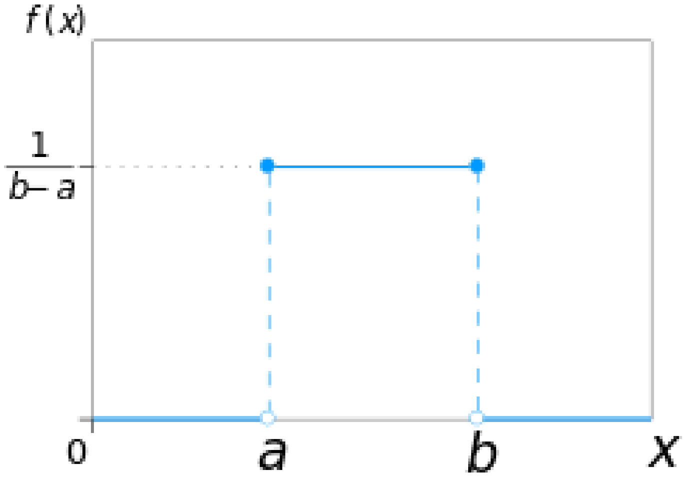
  Poisson regression is the analysis of the relationship between an observed count and a set of explanatory variables (Z) by assuming $\lambda = \alpha + \beta Z$.

CONTENTS

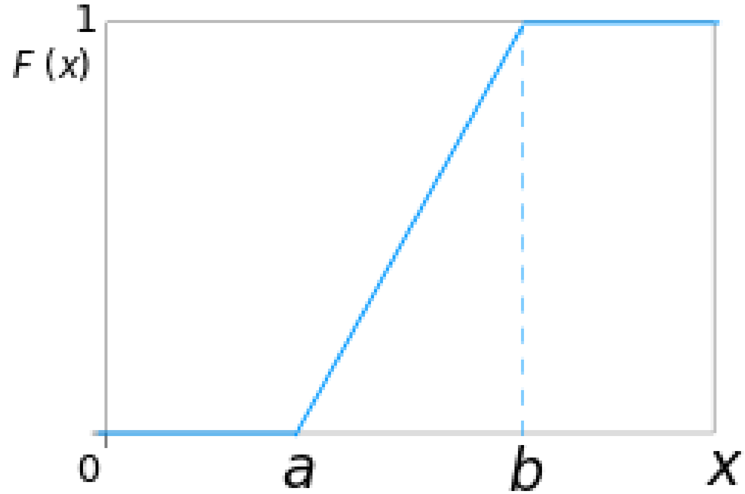
Uniform Distribution (continuous)

A CRV X follows a uniform probability distribution on the interval $[a, b]$, denoted as $X \sim U[a, b]$, if its PDF

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



Probability density function



Cumulative distribution function

Uniform Distribution

Remarks:

- Because X is a bounded random variable, all its moments exist.
- The k-th moment

$$\begin{aligned} E(X^k) &= \int_{-\infty}^{\infty} x^k f_X(x)dx \\ &= \frac{1}{b-a}\int_a^b x^k dx \\ &= \frac{1}{b-a}\frac{x^{k+1}}{k+1}\Big|_a^b \\ &= \frac{1}{b-a}\frac{b^{k+1}-a^{k+1}}{k+1}. \end{aligned}$$

To be Continued

Uniform Distribution

- When $k = 1$, we obtain the mean of X,

$$\mu_X = \frac{1}{2}(a+b).$$

- When $k = 2$, we obtain the second moment

$$E(X^2) = \frac{1}{3}(b^2 + a^2 + ab).$$

- It follows that

$$\begin{aligned} \sigma_X^2 &= E(X^2) - \mu_X^2 \\ &= \frac{1}{12}(b-a)^2. \end{aligned}$$

To be Continued

Uniform Distribution

- The MGF

$$
\begin{aligned}
M_X(t) &= \int_{-\infty}^{+\infty} e^{tx} f_X(x)dx \\
&= \int_a^b e^{tx} \frac{1}{b-a} dx \\
&= \frac{1}{t(b-a)} e^{tx} \Big|_a^b \\
&= \frac{1}{t(b-a)} (e^{tb} - e^{ta}), -\infty < t < \infty.
\end{aligned}
$$

To be Continued

Uniform Distribution

- When $a = 0, b = 1$, the distribution is called the standard uniform U[0,1] distribution, which has the mean $\frac{1}{2}$ and variance $\frac{1}{12}$.
- Recall that the probability integral transform $Y = F_X(X)$ follows a U[0,1] distribution. The uniform distribution plays a very important role in statistics and econometrics.

Beta Distribution

A CRV X follows $Beta(\alpha, \beta)$ if it has a PDF

$$f_X(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 \le x \le 1,$$

where $\alpha > 0, \beta > 0$, and $B(\alpha, \beta)$ is called the Beta function defined as

$$\begin{aligned} \mathrm{B}(\alpha,\beta) &= \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \end{aligned}$$

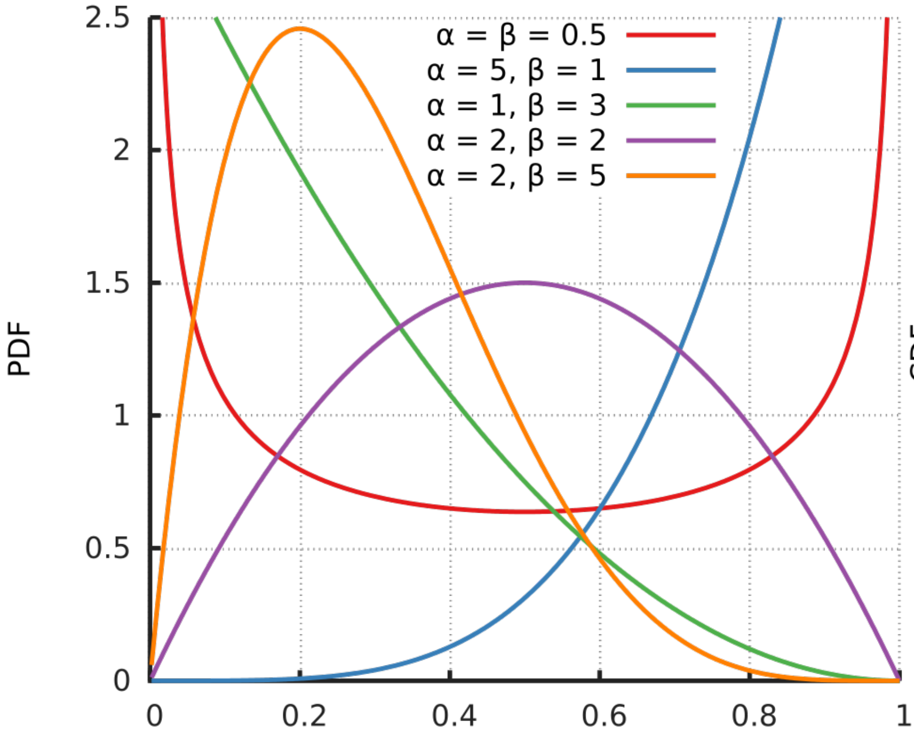
Beta Distribution

Remark:

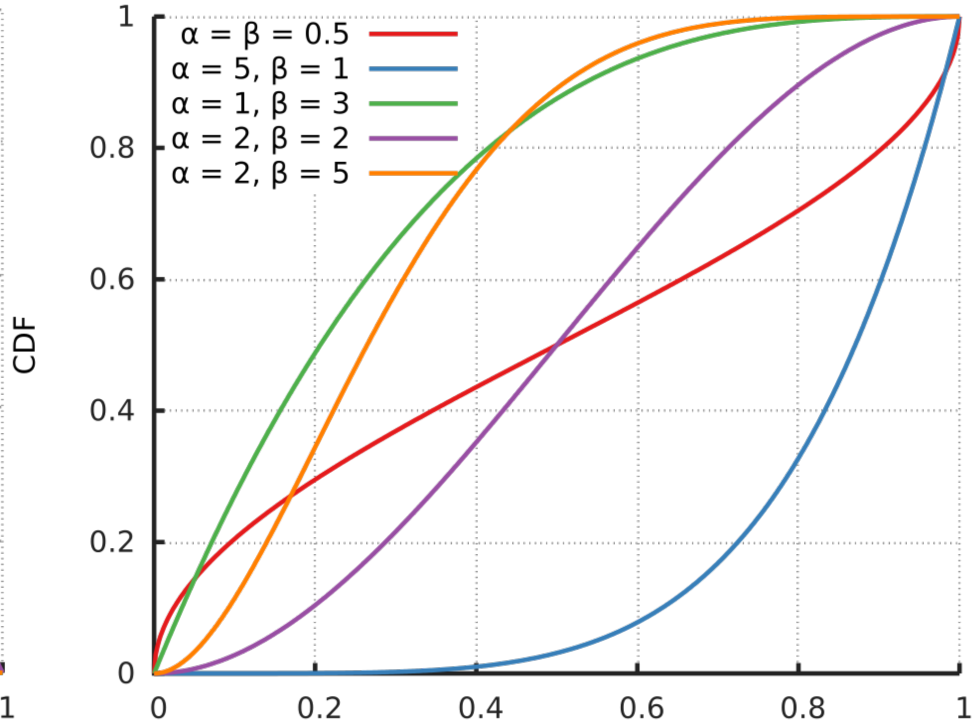
- $\Gamma(\alpha)$ is called the Gamma function, defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

Probability density function



Cumulative distribution function



Beta Distribution

Lemma 1 (3.1). [**Properties of $\Gamma(\alpha)$**]

(1) $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$;

(2) $\Gamma(k) = (k - 1)!$ if k is a positive integer;

(3) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Beta Distribution

Question: How to show the identity that

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}?$$

- Moments of $Beta(\alpha, \beta)$:
  - The mean

$$\begin{aligned} E(X) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{(\alpha+1)-1}(1-x)^{\beta-1}dx \\ &= \frac{\alpha}{\alpha+\beta}\int_0^1 \frac{(\alpha+\beta)\Gamma(\alpha+\beta)}{\alpha\Gamma(\alpha)\Gamma(\beta)} x^{(\alpha+1)-1}(1-x)^{\beta-1}dx \\ &= \frac{\alpha}{\alpha+\beta}\int_0^1 \frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{(\alpha+1)-1}(1-x)^{\beta-1}dx. \\ &= \frac{\alpha}{\alpha+\beta}, \end{aligned}$$

To be Continued

Beta Distribution

– The second moment:

$$
\begin{aligned}
E(X^2) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{(\alpha+2)-1}(1-x)^{\beta-1}dx \\
&= \frac{\alpha}{\alpha+\beta}\int_0^1 \frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{(\alpha+2)-1}(1-x)^{\beta-1}dx \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}\int_0^1 \frac{\Gamma(\alpha+2+\beta)}{\Gamma(\alpha+2)\Gamma(\beta)} x^{(\alpha+2)-1}(1-x)^{\beta-1}dx \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}.
\end{aligned}
$$

To be Continued

Beta Distribution

– The variance

$$\begin{aligned}\text{var}(X) &= E(X^2) - E^2(X)\\ &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - (\frac{\alpha}{\alpha+\beta})^2\\ &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.\end{aligned}$$

Beta Distribution

- The MGF

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} e^{tx} x^{\alpha-1}(1-x)^{\beta-1} dx \\
&= 1 + \sum_{j=1}^{\infty} \left(\prod_{i=0}^{j-1} \frac{\alpha+i}{\alpha+\beta+i} \right) \frac{t^j}{j!},
\end{aligned}
$$

where we have made use of MacLaurin's series expansion

$$
e^{tx} = \sum_{j=0}^{\infty} \frac{t^j x^j}{j!}.
$$

Beta Distribution

- Like the uniform distribution, $Beta(\alpha, \beta)$ is one of the few well-known distributions that have bounded support.
  - $U[0, 1]$, is a special case of $Beta(1, 1)$.
  - The shape of $Beta(\alpha, \beta)$ depends on the values of parameters (α, β). Both α and β are called shape parameters.

Beta Distribution

- Because the support of $Beta(\alpha, \beta)$ is [0,1], the Beta distribution can be used to model the probability distribution of proportions or quantities whose values fall into the interval $[0, 1]$.
  - Granger (1980) uses the Beta distribution for the marginal propensities to consume for individual consumers and show that the sum of "short memory" individual consumption time series displays a "long memory" property the remote past consumption is still persistently correlated with the current consumption.

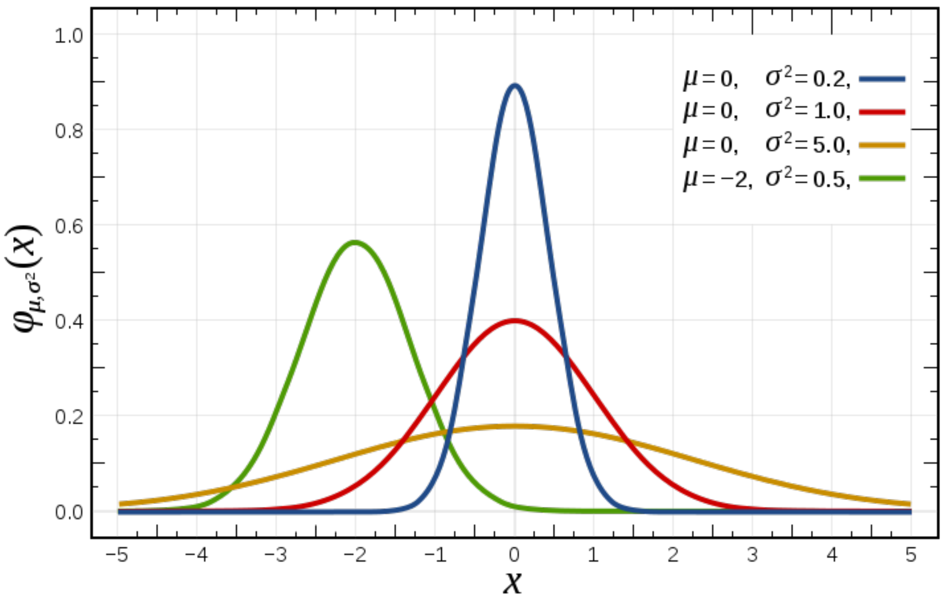
Normal Distribution

A CRV X is called to follow a normal distribution, denoted as $N(\mu, \sigma^2)$, if it has the PDF

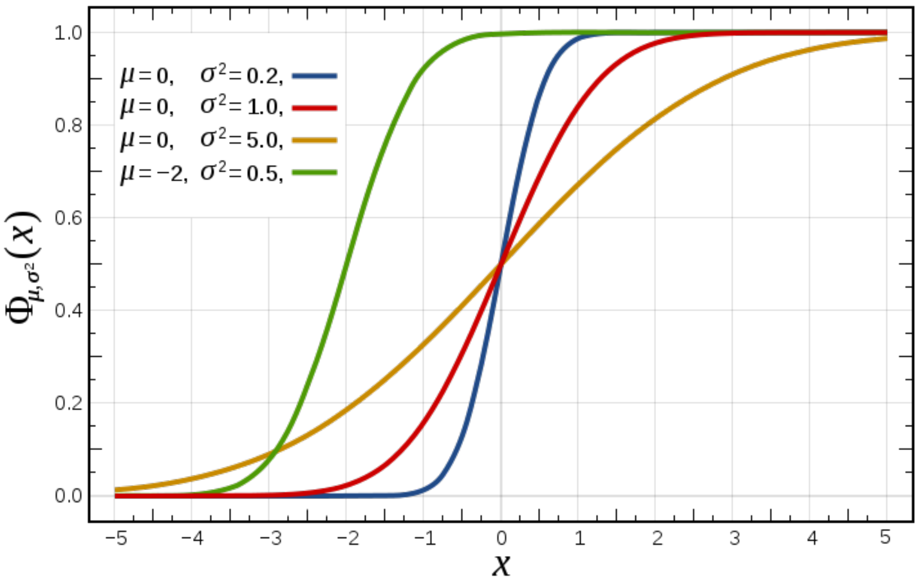
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad -\infty < x < \infty,$$

where $-\infty < \mu < \infty$ and $\sigma^2 > 0$.

Probability density function



Cumulative distribution function



Normal Distribution

Remarks:

- μ and σ^2 are location and scale parameters respectively.
- When $\mu = 0, \sigma^2 = 1$, $N(0, 1)$ is called a standard normal or unit normal distribution.
- The normal distribution was discovered in 1733 by Abraham De Moivre (1667-1754) in his investigation of approximating coin tossing probabilities (i.e., Bernoulli trials with $p = \frac{1}{2}$). He named the PDF of his discovery the exponentially bell-shaped curve.

Normal Distribution

Remarks:

- In 1809, Carl Friedrich Gauss (1777-1855) firmly established the importance of the normal distribution by using it to predict the location of astronomical bodies. As a result, the normal distribution then became commonly known as the Gaussian distribution.

- The normal distribution is perhaps the most important distribution in probability theory. For many decades the normal distribution have been holding a central position in statistics.

Normal Distribution

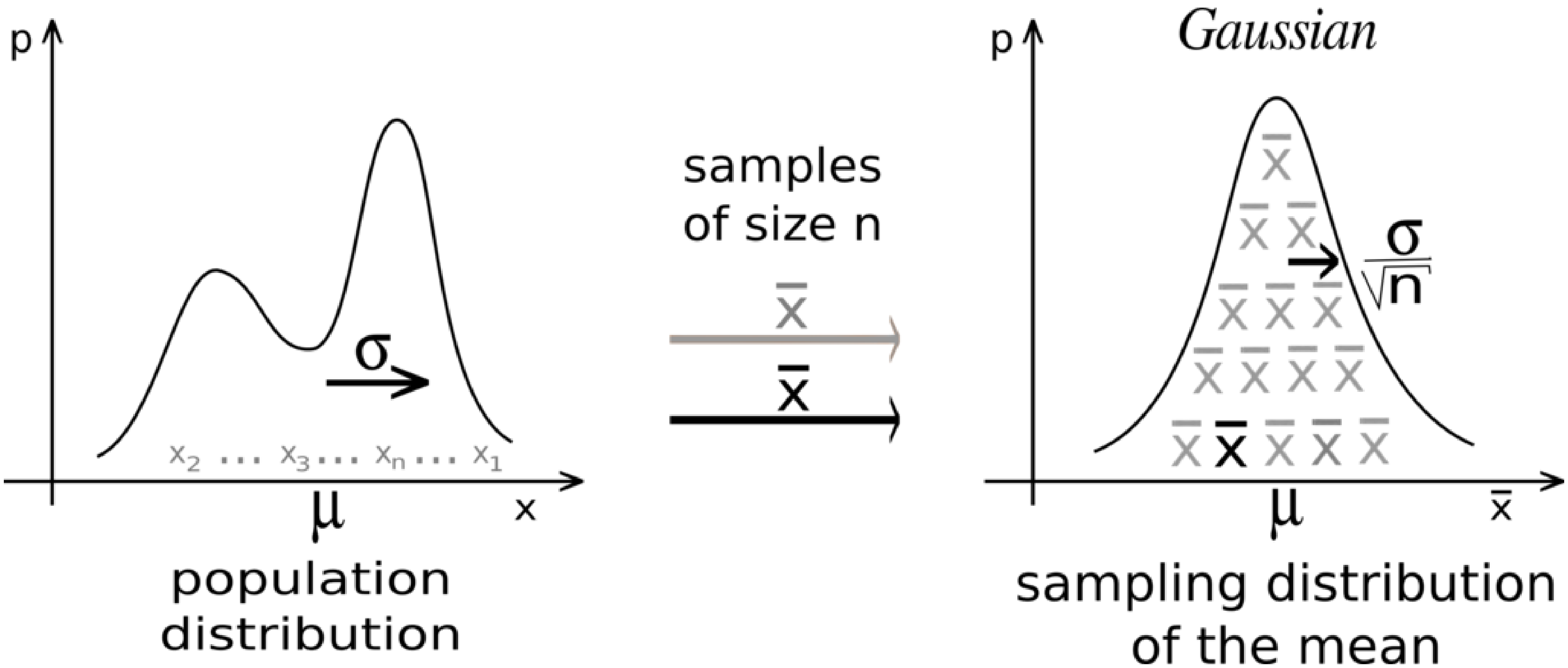
Remarks:

- Most theoretical arguments for the use of the normal distribution are based on the Central Limit Theorem (CLT), which says that under a set of regulrity conditions, a sample average of n IID random variables $\{X_1, \cdots, X_n\}$, with suitable centering and standardization, will converge in distribution to $N(0,1)$ as the sample size n increases:

$$\sqrt{n}\frac{\bar{X}_n - \mu_X}{\sigma_X} \to^d N(0,1) \text{ as } n \to \infty,$$

where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, and the notation $\to^d$denotes convergence in distribution.

Whatever the form of the population distribution, the sampling distribution tends to a Gaussian, and its dispersion is given by the Central Limit Theorem



Normal Distribution

Remarks:

- CLT holds no matter whether X_i is discrete or continuous, and whether X_i has a compactly or uncompact support.

- For example, suppose X_i is a Bernoulli(p) random variable. Then $X = \sum_{i=1}^{n} X_i$ follows a binomial $B(n, p)$ distribution, which can be approximated arbitrarily well by a normal distribution in the sense of

$$\frac{X - np}{\sqrt{np(1-p)}} \to^d N(0, 1) \text{ as } n \to \infty.$$

This was first shown by De Moivre (1718) for $p = \frac{1}{2}$ and then generalized by Laplace (1812) for a general $0 < p < 1$.

Normal Distribution

Question: How to verify

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 \text{ for all } (\mu, \sigma^2)?$$

- The mean

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx + \mu \\
&= \int_{-\infty}^{\infty} (x-\mu) f_X(x) dx + \mu \int_{-\infty}^{\infty} f_X(x) dx \\
&= \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu \qquad \text{(setting } y = x - \mu\text{)} \\
&= \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy + \mu \\
&= \mu.
\end{aligned}$$

To be Continued

Normal Distribution

- The variance

$$
\begin{aligned}
\sigma_X^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \qquad (\text{setting } y = x-\mu) \\
&= \int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{y^2}{2\sigma^2}} dy \\
&= -\sigma^2 \int_{-\infty}^{\infty} yd\left(\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{y^2}{2\sigma^2}}\right) \qquad \left(\text{setting } u = y, v = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2}\right) \\
&= -\sigma^2 \left(y \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{y^2}{2\sigma^2}} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{y^2}{2\sigma^2}} dy \right) \\
&= \sigma^2.
\end{aligned}
$$

Normal Distribution

Theorem 1 (4.2)

Suppose $X \sim N(\mu, \sigma^2)$. Then

$$M_X(t) = e^{\mu t + \frac{\sigma^2}{2} t^2}, \qquad -\infty < t < \infty.$$

Normal Distribution

Proof:

[Method 1]: For the first method, we have

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{tx}e^{-\frac{1}{2\sigma^2}[x^2-2\mu x+\mu^2]}dx \\
&= e^{-\frac{\mu^2}{2\sigma^2}}\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[x^2-2(\mu+\sigma^2 t)x+(\mu+\sigma^2 t)^2-(\mu+\sigma^2 t)^2]}dx \\
&= e^{-\frac{\mu^2}{2\sigma^2}}e^{\frac{(\mu+\sigma^2 t)^2}{2\sigma^2}}\left\{\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{[x-(\mu+\sigma^2 t)]^2}{2\sigma^2}}dx\right\} \\
&= e^{\frac{2\mu\sigma^2 t+\sigma^4 t^2}{2\sigma^2}} \times 1 \\
&= e^{\mu t+\frac{1}{2}\sigma^2 t^2}, \quad \text{for } t \in (-\infty, \infty).
\end{aligned}$$

To be Continued

Normal Distribution

[Method 2]: We note that $X = \mu + \sigma Y$, where $Y \sim N(0,1)$. Then we have

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) \\ &= E\left[e^{t(\mu+\sigma Y)}\right] \\ &= e^{\mu t} E(e^{\sigma t Y}) \\ &= e^{\mu t} M_Y(\sigma t). \end{aligned}$$

To be Continued

Normal Distribution

It suffices to find $M_Y(t)$:

$$\begin{aligned} M_Y(t) &= E\left(e^{tY}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty} e^{-\frac{1}{2}y^2} dy \\ &= e^{\frac{1}{2}t^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-t)^2} dy \\ &= e^{\frac{1}{2}t^2}. \end{aligned}$$

It follows from Theorem 3.20 that

$$\begin{aligned} M_X(t) &= e^{\mu t} M_Y(\sigma t) \\ &= e^{\mu t + \frac{1}{2}\sigma^2 t^2}. \end{aligned}$$

To be Continued

Normal Distribution

- All centered odd moments $E(X-\mu)^{2k+1} = 0$ for all integers $k \geq 0$ because the normal distribution is symmetric about μ.

- Moments $E(X-\mu)^{2k}$:

  One can of course differentiate $M_X(t)$ up to $2k$ times, but this is rather tedious when k is large. We now consider some techniques to calculate the moments $E(X-\mu)^{2k}$ by exploiting the duality between integration and differentiation to obtain the higher order moments of a normal random variable X.

To be Continued

Normal Distribution

– For the first method, put $\beta = \frac{1}{2\sigma^2}$ or $\sigma = \frac{1}{\sqrt{2\beta}}$, we obtain

$$
\begin{aligned}
E(X-\mu)^{2k} &= \int_{-\infty}^{\infty} (x-\mu)^{2k} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} y^{2k} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} y^{2k} e^{-\beta y^2} dy \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (-1)^k \frac{d^k}{d\beta^k} e^{-\beta y^2} dy \\
&= \frac{1}{\sqrt{2\pi}\sigma} (-1)^k \frac{d^k}{d\beta^k} \int_{-\infty}^{\infty} \sqrt{2\pi\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} y^2} dy
\end{aligned}
$$

To be Continued

Normal Distribution

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma}(-1)^k \frac{d^k}{d\beta^k}\left(\sqrt{2\pi}\sigma\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma}(-1)^k \sqrt{2\pi}\frac{d^k}{d\beta^k}\left(\frac{1}{\sqrt{2\beta}}\right) \quad (\text{noting } \sigma = \frac{1}{\sqrt{2\beta}}) \\
&= \frac{1}{\sqrt{2}\sigma}(-1)^k \frac{d^k}{d\beta^k}\left(\beta^{-\frac{1}{2}}\right) \\
&= \frac{1}{\sqrt{2}\sigma}(-1)^k \left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(\frac{1}{2}-k\right)\beta^{-\frac{1}{2}-k} \\
&= \frac{1}{\sqrt{\pi}}\Gamma\left(k+\frac{1}{2}\right)2^k\sigma^{2k},
\end{aligned}
$$

where the last equality is obtained by using $\beta = \frac{1}{2\sigma^2}$ and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

To be Continued

Normal Distribution

– For the special case of $k = 2$, we have

$$E(X-\mu)^4 = (-1)^2 \left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right) 4\sigma^4 = 3\sigma^4.$$

It follows that the kurtosis of $N(\mu, \sigma^2)$ is

$$K = \frac{E(X-\mu)^4}{\sigma^4} = 3.$$

- An alternative approach

Normal Distribution

Lemma 2 (4.3). [**Stein's Lemma**]

Suppose $X \sim N(\mu, \sigma^2)$, and $g(\cdot)$ is a differentiable function satisfying $E|g'(X)| < \infty$. Then

$$E\left[g(X)(X-\mu)\right] = \sigma^2 E\left[g'(X)\right].$$

Normal Distribution

Lemma 2 (4.3). [**Stein's Lemma**]

As an example, let us apply this lemma to calculate $E(X-\mu)^4$. We write

$$\begin{aligned} E(X-\mu)^4 &= E\left[(X-\mu)^3(X-\mu)\right] \\ &= E\left[g(X)(X-\mu)\right], \end{aligned}$$

where $g(X)=(X-\mu)^3$. By Stein's lemma, we have

$$\begin{aligned} E(X-\mu)^4 &= \sigma^2 E\left[3(X-\mu)^2\right] \\ &= 3\sigma^4. \end{aligned}$$

Cauchy and Stable Distributions

A CRV X follows a Cauchy(μ, σ) distribution if its PDF

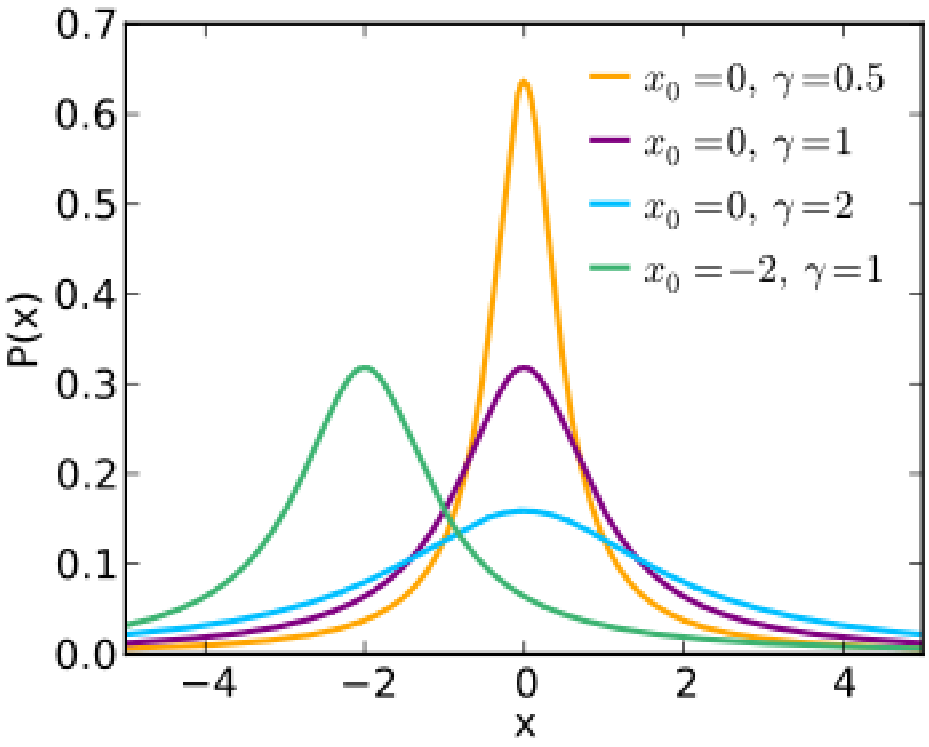
$$f_X(x) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for } -\infty < x < \infty,$$

where $\sigma > 0$.

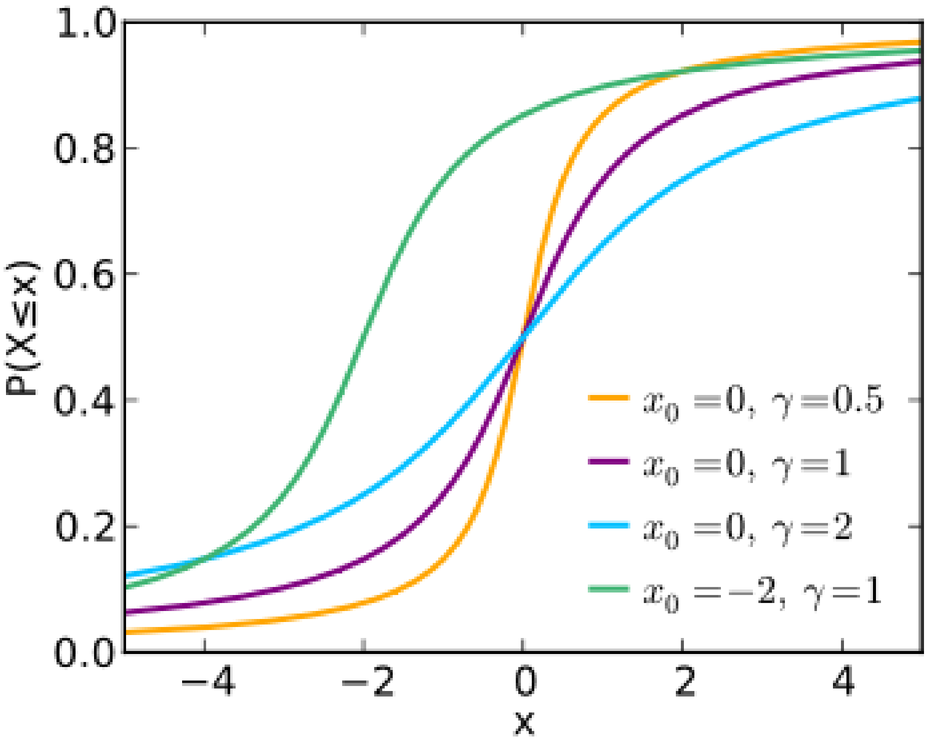
Cauchy and Stable Distributions

- μ and σ are location and scale parameters respectively.
- This distribution is symmetric about μ, with unbounded support. When $\mu = 0$ and $\sigma = 1$, the distribution is called a standard Cauchy distribution, denoted as Cauchy$(0, 1)$.
- There has been little use of the Cauchy distribution in practice. However, it is of special theoretical importance. In particular, the Cauchy distribution has some peculiar properties and could provide counter examples to some generally accepted results and concepts in statistics.

Probability density function



Cumulative distribution function



Cauchy and Stable Distributions

- The Cauchy distribution has very long and heavy tails. The most notable difference between the normal and Cauchy distributions is in the flatter tails of the latter.

- For Cauchy(μ, σ), the tail of the PDF decays to zero at a very slow hyperbolic rate: $f_X(x) \sim x^{-2}$ as $|x| \to \infty$. As a consequence, all finite moments of order greater than or equal to 1 do not exist, and so its MGF does not exist either. This implies the mean and all higher order moments do not exist.

Cauchy and Stable Distributions

- The location parameter μ cannot be interpreted as the mean, and the scale parameter σ cannot be interpreted as the standard deviation.

- The characteristic function of Cauchy(μ, σ) is

$$\begin{aligned} \varphi_X(t) &= E\left(e^{\mathbf{i}tX}\right) \\ &= e^{\mathbf{i}\mu t - \sigma|t|}. \end{aligned}$$

  This characteristic function is not differentiable with respect to t at the origin, which is consistent with the fact that all finite moments do not exist.

Cauchy and Stable Distributions

Question: When can a Cauchy distribution arise?

- A Cauchy distribution can arise when a ratio of two independent normal random variables is considered.
- Both the Cauchy and normal distributions belong to the class of so-called stable distributions, which has very considerable importance in probability theory.

Cauchy and Stable Distributions

Question: What is a stable distribution?

- For a stable distribution, its PDF is usually unknown in closed form. However, its characteristic function has a closed form

$$\varphi_X(t) = e^{\mathbf{i}\mu t - \sigma|t|^c[1+\mathbf{i}\lambda\ \mathrm{sgn}(t)\omega(|t|,c)]},$$

where $\mathbf{i} = \sqrt{-1}$, $0 < c \leq 2, -1 \leq \lambda \leq 1, \sigma > 0$, and

$$\omega(|t|, c) = \begin{cases} \tan\left(\frac{1}{2}\pi c\right), & c \neq 1, \\ -\frac{2}{\pi}\ln(|t|), & c = 1, \end{cases}$$

and

$$\mathrm{sgn}(t) = \begin{cases} 1, & \text{if } t > 0, \\ 0, & \text{if } t = 0, \\ -1, & \text{if } t < 0. \end{cases}$$

To be Continued

Cauchy and Stable Distributions

- Intuitively, μ is a location parameter, σ is a scale parameter, c is a tail parameter, and λ is a skew parameter.
  - The shape of the PDF is determined by both c and λ.
  - If $\lambda = 0$, the distribution is symmetric.
  - When $\lambda = 0$, $c = 2$ gives a normal distribution, and
  - when $\lambda = 0, c = 1$ gives a Cauchy distribution. meters $(\mu, \sigma, c, \lambda)$.
- The moments of a stable distribution exist only when $c > 1$.

To be Continued

Cauchy and Stable Distributions

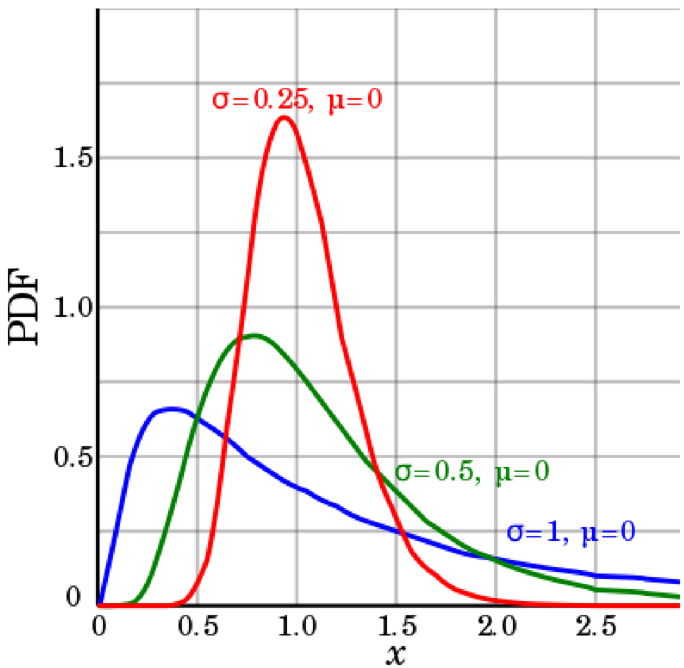
- The sum of stable random variables has a limiting distribution, which is a stable distribution rather than a normal distribution. Thus nonnormal stable distributions generalize the Central Limit Theorem (CLT) to the cases where the second moment of the summed variables are infinite.

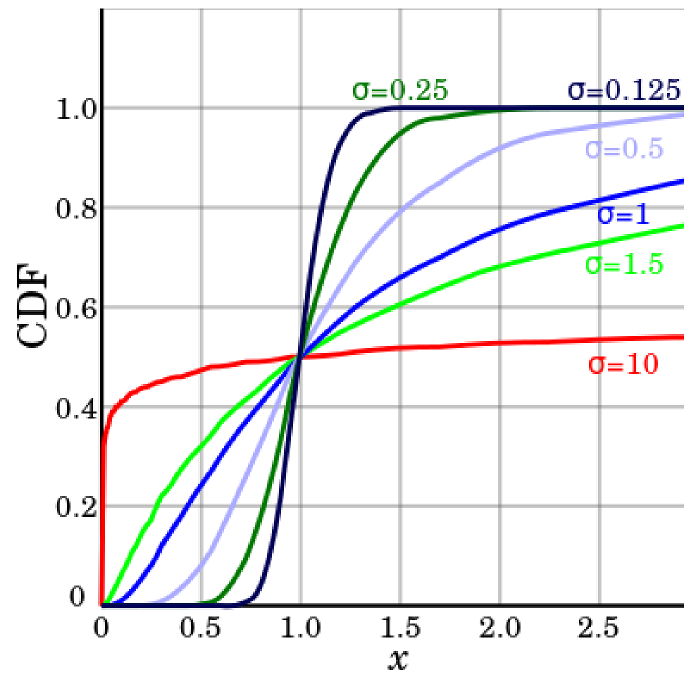
- The stable distributions are closely related to the Levy process which has recently become a popular tool in financial econometrics. They are more appropriate to model heavy tails which are often observed in financial data. Mandelbrot (1963) and Fama (1965) have applied the stable distributions to model stock returns.

Lognormal Distribution

A CRV X follows a Lognormal distribution, denoted as $LN(\mu, \sigma^2)$, if its PDF

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}e^{\frac{1}{2\sigma^2}(\ln x - \mu)^2}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$





Lognormal Distribution

Remarks:

- $Y = \ln(X) \sim N(\mu, \sigma^2)$. Indeed, X is called a lognormal random variable because its logarithm follows a normal distribution.
- The lognormal distribution is sometimes called the antilognormal distribution. This name has some logical basis in that it is not the distribution of the logarithm of a normal variable but of the exponential—antilogarithmic function of such a variable.
- There is a more general definition of the lognormal distribution. Suppose $Y = \ln(X - \alpha) \sim N(\mu, \sigma^2)$. Then X is said to follow a lognormal distribution $LN(\alpha, \mu, \sigma^2)$. Since parameter α only affects the mean of X, we consider the two-parameter $LN(\mu, \sigma^2)$ distribution here.

Lognormal Distribution

- Since the MGF of a normal random variable $Y \sim N(\mu, \sigma^2)$ is

$$M_Y(t) = E\left(e^{tY}\right) = e^{\mu t + \frac{\sigma^2}{2}t^2},$$

it follows that all moments of the lognormal(μ, σ^2) random variable X exist and are given by

$$\begin{aligned} E(X^k) &= E\left(e^{kY}\right) \\ &= M_Y(k) \\ &= e^{k\mu + \frac{\sigma^2}{2}k^2}, \qquad k = 1, 2, \cdots. \end{aligned}$$

To be Continued

Lognormal Distribution

In particular, we have the mean

$$\mu_X = e^{\mu + \frac{\sigma^2}{2}},$$

and the variance

$$\begin{aligned} \sigma_X^2 &= E(X^2) - \mu_X^2 \\ &= e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right). \end{aligned}$$

- It is important to note that parameters μ and σ^2 are not the mean and variance of the $LN(\mu, \sigma^2)$ distribution.

To be Continued

Lognormal Distribution

- Although all moments exist, the MGF does not exist for a lognormal distribution. To see this, we consider

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \int_0^\infty e^{tx} \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{1}{2\sigma^2}(\ln x-\mu)^2} dx \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{te^{\ln x}-\frac{1}{2\sigma^2}(\ln x-\mu)^2} d\ln x \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{te^y-\frac{1}{2\sigma^2}(y-\mu)^2} dy \text{ (by setting } y=\ln x) \\
&\geq \int_c^{c+1} \frac{1}{\sqrt{2\pi}\sigma} e^{te^y-\frac{1}{2\sigma^2}(y-\mu)^2} dy \text{ for any } c>0 \\
&\geq \frac{1}{\sqrt{2\pi}\sigma} e^{te^c-\frac{1}{2\sigma^2}(c-\mu)^2}(c+1-c) \text{ for } c \text{ sufficiently large} \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{te^c-\frac{1}{2\sigma^2}(c-\mu)^2} \\
&\to \infty \text{ as } c \to \infty
\end{aligned}
$$

because $te^c - (c-\mu)^2/2\sigma^2 \to \infty$ as $c \to \infty$ and $t > 0$.

Lognormal Distribution

Question: What is the characteristic function of the $LN(\mu, \sigma^2)$ distribution?

- The lognormal distribution is very popular in modeling applications when the variable of interest is nonnegative and skewed to the right. In particular, it has been widely used to model the distribution of asset prices, commodity prices, incomes and populations.
  - Example of stochastic compound growth: consider following nonnegative economic variable

$$X_t = X_{t-1}(1 + Y_t),$$

where $\{Y_t\}$ are a sequence of IID random variables such that Y_t is independent of X_{t-1}.

To be Continued

Lognormal Distribution

– The recursive relationship leads to

$$X_n = X_0 \prod_{t=1}^{n} (1 + Y_t)$$

and so

$$\ln X_n = \ln X_0 + \sum_{t=1}^{n} \ln(1 + Y_t).$$

By the CLT, for large n, $\sum_{t=1}^{n} \ln(1 + Y_t)$ will be approximately normally distributed, after suitable standardization. Thus, its exponential, $\Pi_{t=1}^{n} (1 + Y_t)$ is approximately lognormally distributed.

To be Continued

Lognormal Distribution

- The lognormality assumption offers a great deal of convenience in analysis. Suppose a stock price $P_t \sim LN(\mu t, \sigma^2 t)$, where the time t changes continuously. Then $\ln P_t \sim N(\mu t, \sigma^2 t)$, and the log-return

$$\begin{aligned} R_t &= \ln(P_t/P_{t-1}) \\ &= \ln(P_t) - \ln(P_{t-1}) \end{aligned}$$

which is approximately equal to the relative price change from time $t-1$ to time t, is also normally distributed. Furthermore, the cumulative return $\sum_{t=1}^{m} R_t$ over m time periods from $t = 1$ to $t = m$ is also normally distributed.

To be Continued

Lognormal Distribution

- Applications:
  - Black and Scholes (1973) use the lognormal distribution for the underlying stock price in deriving the European options prices.
  - Eeckhout (2004), based on Census 2000 data, documented that the size of all cities in the United States follows a lognormal distribution, and proposed an interesting equilibrium model of economic activity with local externality to explain the longnormal distribution.

To be Continued

Lognormal Distribution

- The relationship between leaving a company and employees tenure has been described by lognormal distributions with great success (Young 1971, McClean 1976).
- O'Neill and Wells (1972) point out that the lognormal distribution can be effectively shed to fit the distribution for individual insurance claim payments.

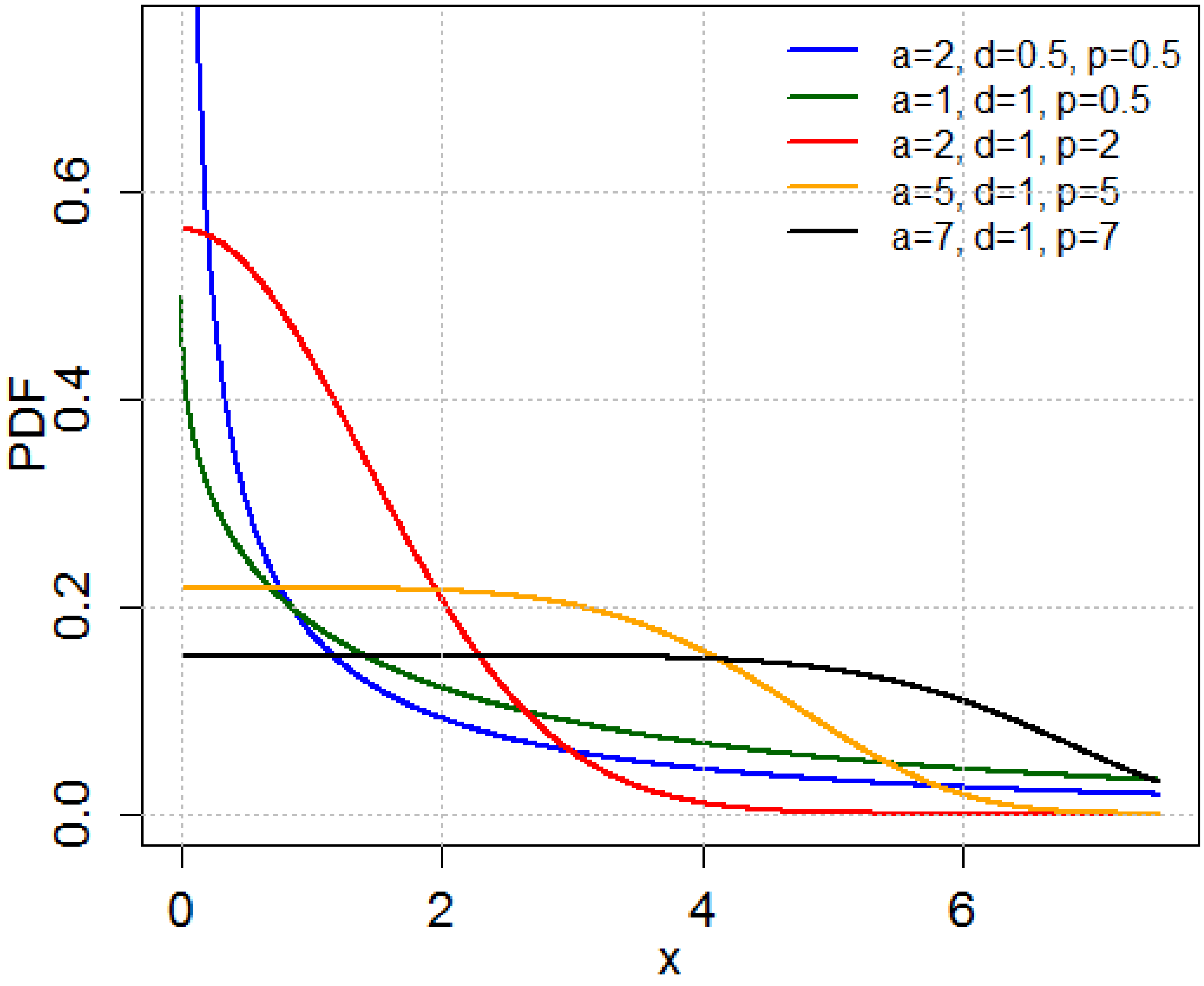
Gamma and Generalized Gamma Distributions

- A nonnegative CRV X follows a Gamma distribution, denoted as $G(\alpha, \beta)$, if its PDF

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

where $\alpha, \beta > 0$, and $\Gamma(\alpha)$ is the Gamma function.

Probability density function



Gamma and Generalized Gamma Distributions

- $G(\alpha, \beta)$ is a flexible family of distributions for a nonnegative random variable on $[0, \infty)$.
  - α is a shape parameter, and β is a scale parameter controlling the spread of the distribution.
  - When $\beta = 1$, $G(\alpha, 1)$ is called a standard Gamma distribution.

Gamma and Generalized Gamma Distributions

Remarks:

- The mean

$$
\begin{aligned}
\mu_X &= \int_{-\infty}^{\infty} x f_X(x) dx \\
&= \int_0^{\infty} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha} e^{-x/\beta} dx \\
&= \int_0^{\infty} \frac{\alpha\beta}{\alpha\Gamma(\alpha)\beta^{(a+1)}} x^{(\alpha+1)-1} e^{-x/\beta} dx \quad \text{setting } \alpha^* = \alpha + 1 \\
&= \alpha\beta \int_0^{\infty} \frac{1}{\Gamma(\alpha^*)\beta^{\alpha^*}} x^{a^*-1} e^{-x/\beta} dx \\
&= \alpha\beta.
\end{aligned}
$$

Gamma and Generalized Gamma Distributions

Remarks:

- The variance

$$\sigma_X^2 = E(X^2) - \mu_X^2 = \alpha\beta^2.$$

Gamma and Generalized Gamma Distributions

Remarks:

- The MGF

$$
\begin{aligned}
M_X(t) &= \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx \\
&= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x(1/\beta - t)} dx \quad \left(\text{setting } \beta^* = \frac{1}{1/\beta - t}\right) \\
&= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta^*} dx \\
&= \frac{(\beta^*)^\alpha}{\beta^\alpha} \int_0^\infty \frac{1}{\Gamma(\alpha)(\beta^*)^\alpha} x^{\alpha-1} e^{-x/\beta^*} dx \\
&= \frac{(\beta^*)^\alpha}{\beta^\alpha} \\
&= (1-\beta t)^{-\alpha}, \quad t < 1/\beta.
\end{aligned}
$$

Gamma and Generalized Gamma Distributions

Remarks:

- Gamma(α, β) has been used to model the distribution of the continuous waiting time of economic events (e.g., unemployment duration, price duration, poverty duration, etc.). It can also be used to model distribution of non-negative random variables, such as income, population, and range.

Gamma and Generalized Gamma Distributions

Remarks:

- Cox, Ingersoll and Ross (1985) propose an equilibrium model for the term structure of the spot interest rate. They assume that the spot interest rate

$$dr_t = k(\theta - r_t)dt + \sigma\sqrt{r_t}dB_t$$

where B_t is a Brownian motion process. For $k, \theta > 0$, the interest rate is elastically pulled toward a central location or long-term value θ, and the parameter k determines the speed of adjustment.

Gamma and Generalized Gamma Distributions

Remarks:

- It can be shown that the Cox, Ingersoll and Ross (1985) model admits a steady state Gamma distribution with PDF:

$$f(r) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} r^{\alpha-1} e^{-r/\beta}$$

where $\alpha = \frac{2k\theta}{\sigma^2}$ and $\beta = \frac{\sigma^2}{2k}$.

Gamma and Generalized Gamma Distributions

- There is a closely related distribution called Generalized Gamma distribution.

  Suppose the random variable

$$Y = \left(\frac{X-\gamma}{\beta}\right)^c$$

  follows a standard Gamma distribution, i.e., $Y \sim \text{Gamma}(\alpha, 1)$ or equivalently it has the PDF

$$f_Y(y) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y}, \qquad y \geq 0.$$

  Then X is said to follow a Generalized Gamma distribution with parameters where α and c are shape parameters, β is a scale parameter, and γ is a location parameter.

Gamma and Generalized Gamma Distributions

- The PDF of X

$$f_X(x) = \frac{c}{\Gamma(\alpha)\beta^{c\alpha}}(x-\gamma)^{c\alpha-1}e^{-\left(\frac{x-\gamma}{\beta}\right)^c}, \qquad x \geq \gamma.$$

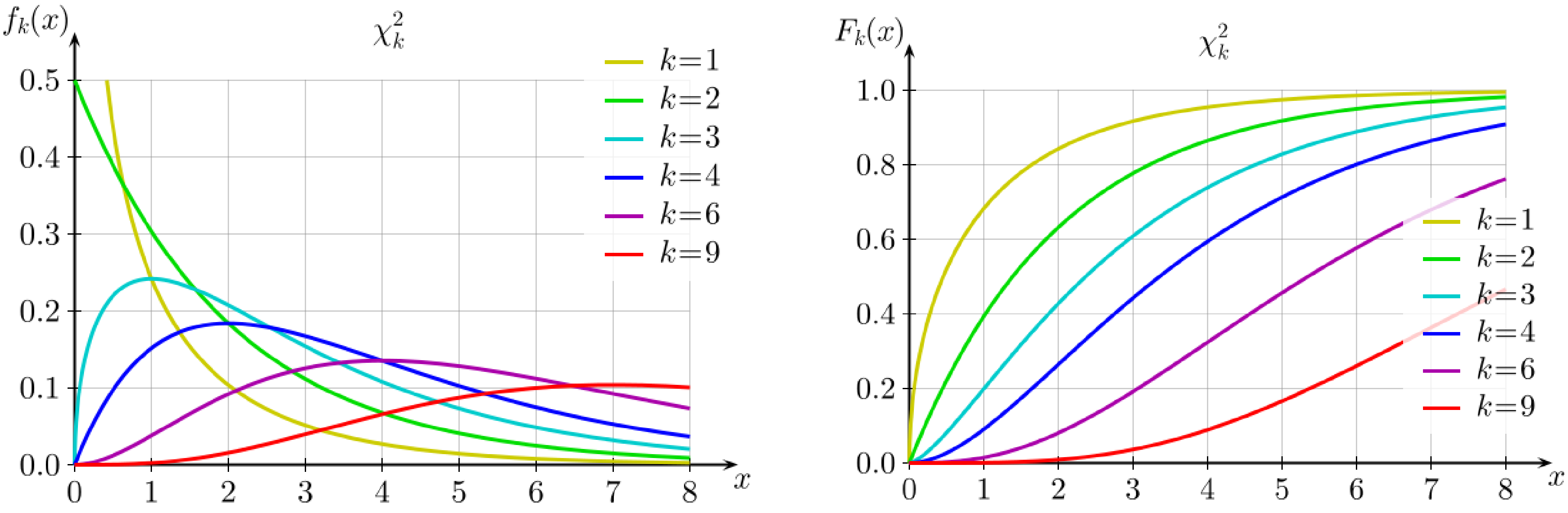
- The moments of the Generalized Gamma distribution can be obtained from the moments of $G(\alpha, 1)$ by observing the following relationship:

$$\begin{aligned} E\left[\left(\frac{X-\gamma}{\beta}\right)^k\right] &= E\left[\left(\frac{X-\gamma}{\beta}\right)^{c(k/c)}\right] \\ &= E[Y^{(k/c)}] \\ &= \frac{\Gamma(\alpha+k/c)}{\Gamma(\alpha)}. \end{aligned}$$

Chi-Square Distribution

A nonnegative CRV X follows a Chi-square distribution with ν degrees of freedom, denoted as χ^2_ν, if its PDF

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\frac{\nu}{2})\sqrt{2^\nu}} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$



Chi-Square Distribution

- χ^2_ν is a special case of $G(\alpha, \beta)$ with $\alpha = \frac{\nu}{2}$, and $\beta = 2$.
- Its k-th moment is

$$E(X^k) = \frac{2^k \Gamma(\frac{\nu}{2} + k)}{\Gamma(\frac{\nu}{2})}.$$

In particular, its mean is

$$E(X) = \nu;$$

its variance is

$$\mathrm{var}(X) = 2\nu;$$

Chi-Square Distribution

- The MGF

$$M_X(t) = (1-2t)^{-\frac{\nu}{2}} \text{ for } t < \frac{1}{2}.$$

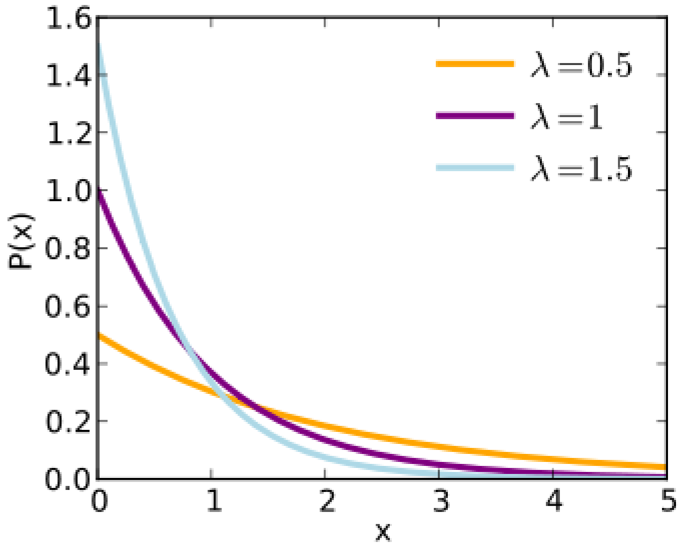
- χ^2_ν allows ν, the degree of freedom parameter, to be a non-integer value. When ν is an integer, χ^2_ν is equivalent to that of the sum of ν squared independent $N(0,1)$ random variables.
- χ^2_ν is a right-skewed distribution. When the degree of freedom $\nu \to \infty$, χ^2_ν becomes an approximately normal distribution with mean ν and variance 2ν.

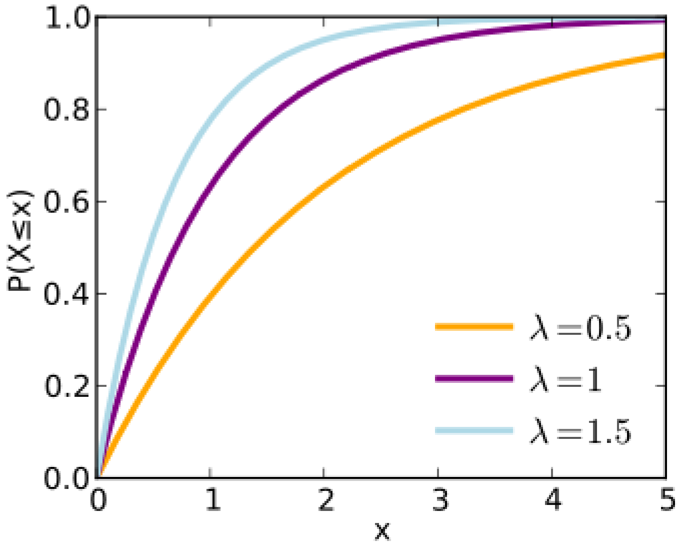
Exponential and Weibull Distributions

A nonnegative CRV X follows an exponential distribution, denoted as $EXP(\beta)$, if its PDF

$$f_X(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

where $\beta > 0$.





Exponential and Weibull Distributions

Remarks:

- β is a scale parameter. When $\beta = 1$, X is called to follow the standard exponential distribution, denoted as $EXP(1)$.
- $EXP(\beta)$ is a special case of $G(1, \beta)$.
- The MGF of X

$$M_X(t) = E(e^{tX}) = \frac{1}{1 - \beta t}, \qquad t < \frac{1}{\beta}.$$

Exponential and Weibull Distributions

Remarks:

- The mean
$$E(X) = \beta.$$
- The variance
$$\operatorname{var}(X) = \beta^2.$$
- Like the geometric distribution, $EXP(\beta)$ also has the so-called "memoryless" property: For any positive numbers x and y, $x > y$,
$$P(X > x | X > y) = P(X > x - y).$$

Exponential and Weibull Distributions

Remarks:

- $EXP(\beta)$ can be viewed as a continuous analog of the discrete geometric distribution. In fact, $EXP(\beta)$ is the only continuous distribution with memoryless property

(**Question**: How to show this?)

Exponential and Weibull Distributions

The Poisson process provides an interesting unified framework to link the exponential and Gamma distributions.

Question: What is the Poisson process?

To be Continued

Exponential and Weibull Distributions

Suppose $\{N(t) : t \geq 0\}$ is a stationary Poisson process with rate λ. Let X_1 be the time of the first event, and for $n \geq 2$, let X_n be the time between the $(n-1)$-th and the n-th events. Then it can be shown that $\{X_1, X_2, ...\}$ is a sequence of IID exponential random variables; that is, $X_i \sim EXP(1/\lambda)$. Next, let $X = \sum_{i=1}^{n} X_i$ be the time of the n-th event. Then it can be shown that X, which is the sum of a sequence of IID exponential random variables, follows $G(n, 1/\lambda)$.

Exponential and Weibull Distributions

Question: Why is the exponential distribution useful in economics and finance?

- The exponential distribution is of considerable importance and widely used in statistics and econometrics. The exponential distribution has been used to model time durations of economic events, such as the unemployment spell of a worker, time before a credit default, the time between two trades or price changes, etc.

To be Continued

Exponential and Weibull Distributions

- There are many situations where one would expect an exponential distribution to give a useful description of observed phenomena.
    - An example in labor economics: Let X be the unemployment duration of a worker which has a PDF $f_X(x)$. Then the so-called hazard rate or hazard function is defined as

$$
\begin{aligned}
\lambda(x) &= \lim_{\Delta x \to 0^+} \frac{P(X \leq x + \Delta x | X \geq x)}{\Delta x} \\
&= \lim_{\Delta x \to 0^+} \frac{P(x \leq X \leq x + \Delta x)}{P(X \geq x)\Delta x} \\
&= \left[\lim_{\Delta x \to 0^+} \frac{\int_x^{x+\Delta x} f_X(u)du}{\Delta x} \right] \cdot \frac{1}{P(X \geq x)}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{f_X(x)}{P(X \geq x)} \\
&= \frac{f_X(x)}{1 - F_X(x)} \\
&= -\frac{d}{dx} \ln[1 - F_X(x)]
\end{aligned}
$$

To be Continued

Exponential and Weibull Distributions

- Intuitively, the hazard function $\lambda(x)$ is the instantaneous probability that the unemployed worker will find a job after an unemployment duration of x. Duration analysis, which is more popularly called survival analysis in statistics, is to model $\lambda(x)$ using economic explanatory variables.
- An example in labor economics: Assume that the hazard rate is a constant function of the unemployment duration x,i.e.,

$$\lambda(x) = \lambda_0 \text{ for all } x.$$

To be Continued

Exponential and Weibull Distributions

– Then the corresponding distribution of the unemployment duration X follows an $EXP(1/\lambda_0)$ distribution:

$$f_X(x) = \lambda_0 e^{-\lambda_0 x} \text{ for } x > 0.$$

– An example in financial econometrics: An empirical stylized fact of high-frequency stock return $\{X_t\}$ is that the absolute value of the stock return $|X_t|$ approximately follows a standard exponential distribution (Ding, Granger and Engle 1993). Here, X_t is the standardized financial return in time period t.

Exponential and Weibull Distributions

Weibull Distribution:

- Suppose $Y = (X - \alpha)^c$ follows an $EXP(\beta)$. Then X is said to have a Weibull distribution. Its PDF is

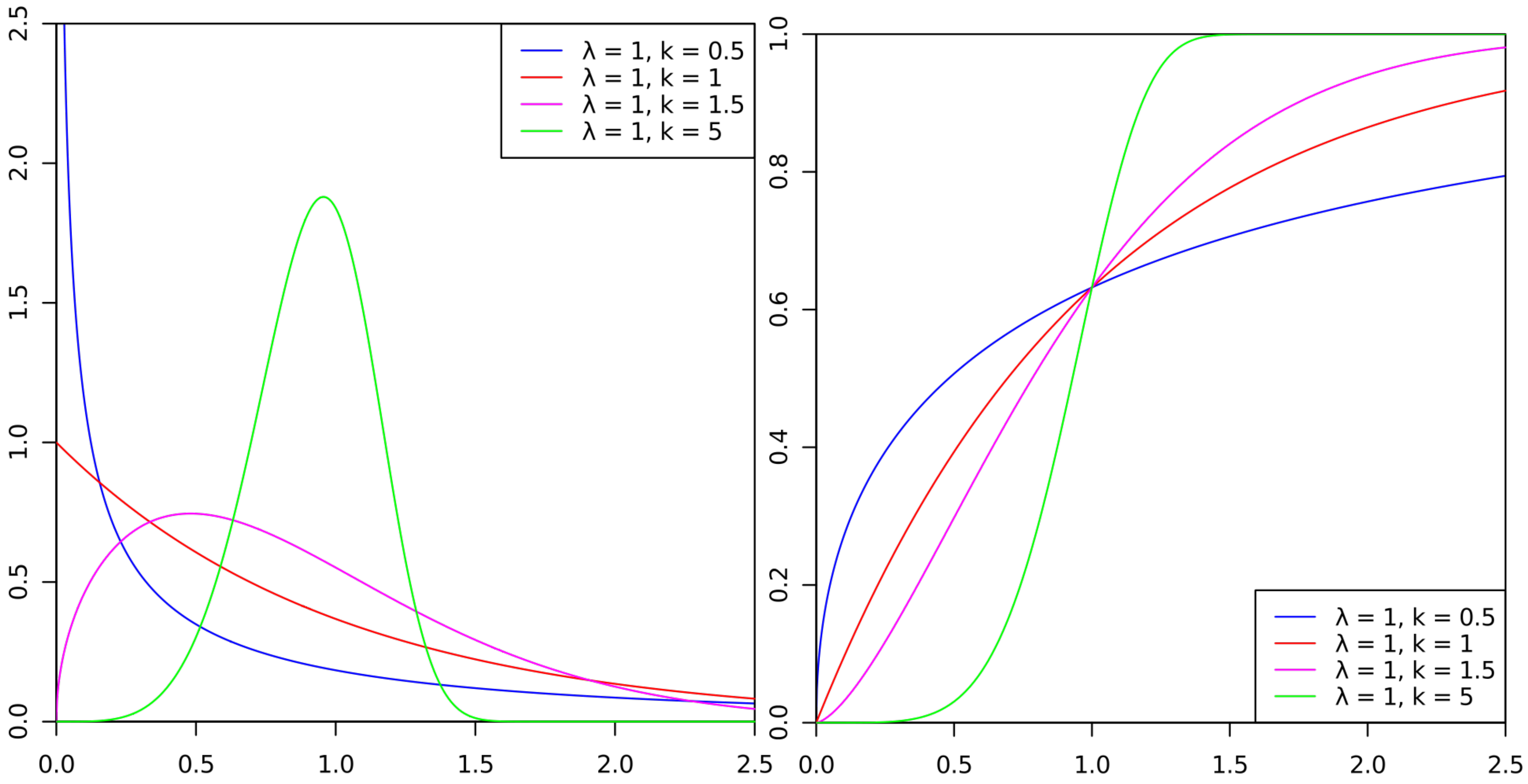
$$f_X(x) = \frac{c}{\beta}(x-\alpha)^{c-1} e^{-\frac{(x-\alpha)^c}{\beta}}, \qquad x > \alpha,$$

  where α is a location parameter, β is a scale parameter, and c is a shape parameter. It is necessary that c be greater than 1 (why?).

- The Weibull distribution is more flexible than the exponential distribution. For example, the associated hazard function is no longer a constant function.

- The Weibull distribution is used in Engle and Russell (1998) to model the time duration between trades or price changes in finance.

Weibull (2-parameter)

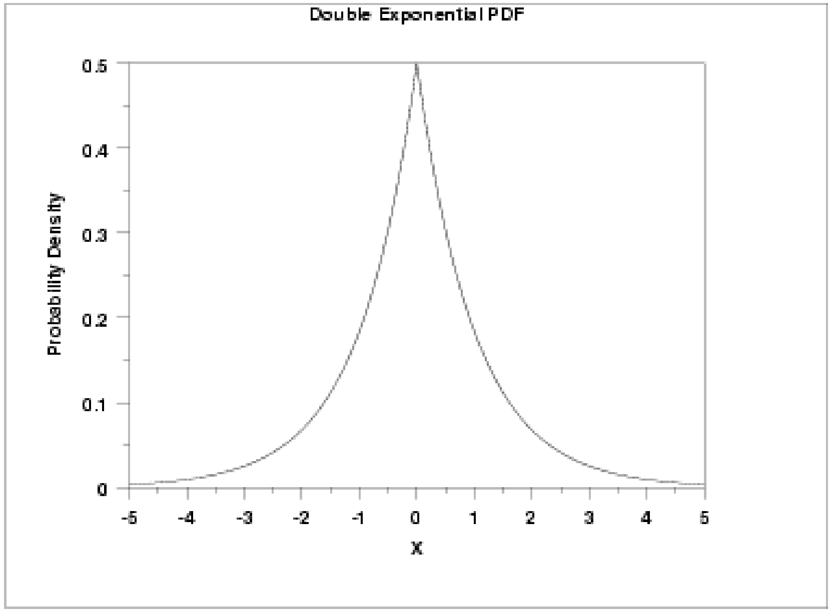


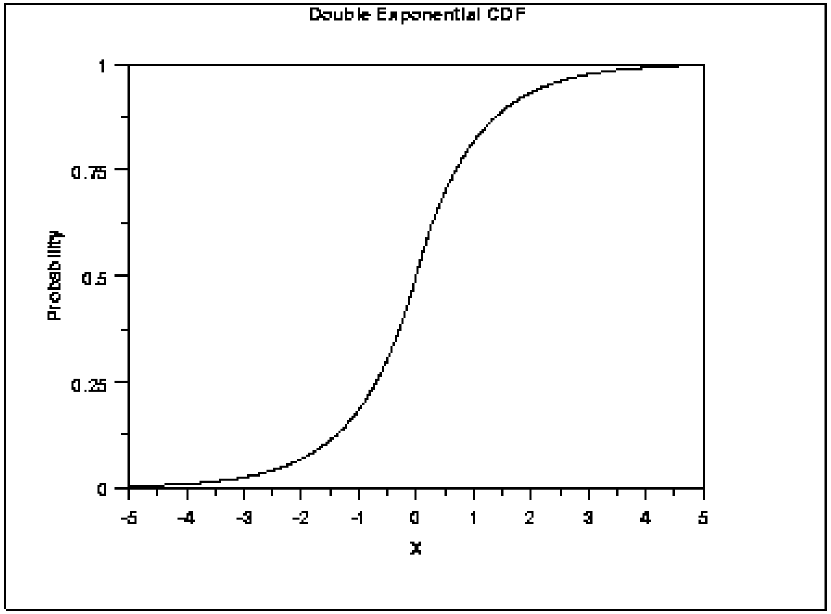
Double Exponential Distribution

A continuous random variable X follows a double exponential distribution, denoted as $DEXP(\alpha, \beta)$, if its PDF

$$f_X(x) = \frac{1}{2\beta} e^{-\frac{|x-\alpha|}{\beta}}, \qquad -\infty < x < \infty,$$

where $\beta > 0$.





Double Exponential Distribution

Remarks:

- The double exponential distribution is a symmetric distribution about α, but has a fatter tail than a normal distribution.
- It has a peak at $x = \alpha$ where the derivative does not exist.
- The mean of X,

$$E(X) = \alpha,$$

Double Exponential Distribution

Remarks:

- the variance

$$\operatorname{var}(X) = 2\beta^2,$$

- The MGF

$$M_X(t) = \frac{e^{\alpha t}}{1 - (\beta t)^2} \text{ for } |t| < \frac{1}{\beta}.$$

- The double exponential distribution is also called the Laplace distribution.

CONTENTS

Conclusion

Conclusion:

- We have introduced a variety of important parametric discrete and continuous probability distributions, and investigated their properties and their relationships with each other.
- The flexibility of these parametric distributions depends on the functional form of the PDF, and the number of parameters. We have paid a closed attention to the interpretation of the parameters and their relationships.
- Many of these distributions related to each other, and they are widely used in modeling economic and financial data.

Thank You !