

# **Credibility Theory and Loss distribution**

Mohamed E. Ghoneim  
Mathematics Department  
Umm Al-Qura University

# Objective of Credibility Theory

The Credibility Theory and Loss Distributions course focuses on:

- ❑ Developing a comprehensive understanding of actuarial models, frequency and severity analysis, and aggregate loss models while applying statistical estimation techniques.
- ❑ It emphasizes on empirical model construction, credibility methods, and loss distribution applications in actuarial practice loss distribution applications in actuarial practice.

| No | List of Topics                                                                                                                                                                                                                                                                                                                               | Contact Hours |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| 1. | Introduction: Modeling, Random variables, Key Functions, and Four Models and Basic Distributional Quantities. Characteristics of Actuarial Models, Creating New Distributions, Selected Distributions, and Their Relationships and The Linear Exponential Family. Frequency and Severity with Coverage Modifications. Aggregate Loss Models. | 15            |
| 2. | Mathematical Statistics: Moments, and Percentile Estimation, Maximum Likelihood Estimation (MLE) and Bayesian Estimation. Construction of Empirical Models and Model Selection.                                                                                                                                                              | 10            |
| 3. | Limited Fluctuation Credibility Theory, Full Credibility and Partial Credibility                                                                                                                                                                                                                                                             | 6             |
| 4. | Greatest Accuracy Credibility: Conditional Distributions and Expectation, The Bayesian Methodology, The Credibility Premium, The Bühlmann Model, The Bühlmann-Straub Model and Exact Credibility.                                                                                                                                            | 10            |
| 5. | General Applications in Credibility Theory and Loss Distributions                                                                                                                                                                                                                                                                            | 4             |

# Claim Distribution

---

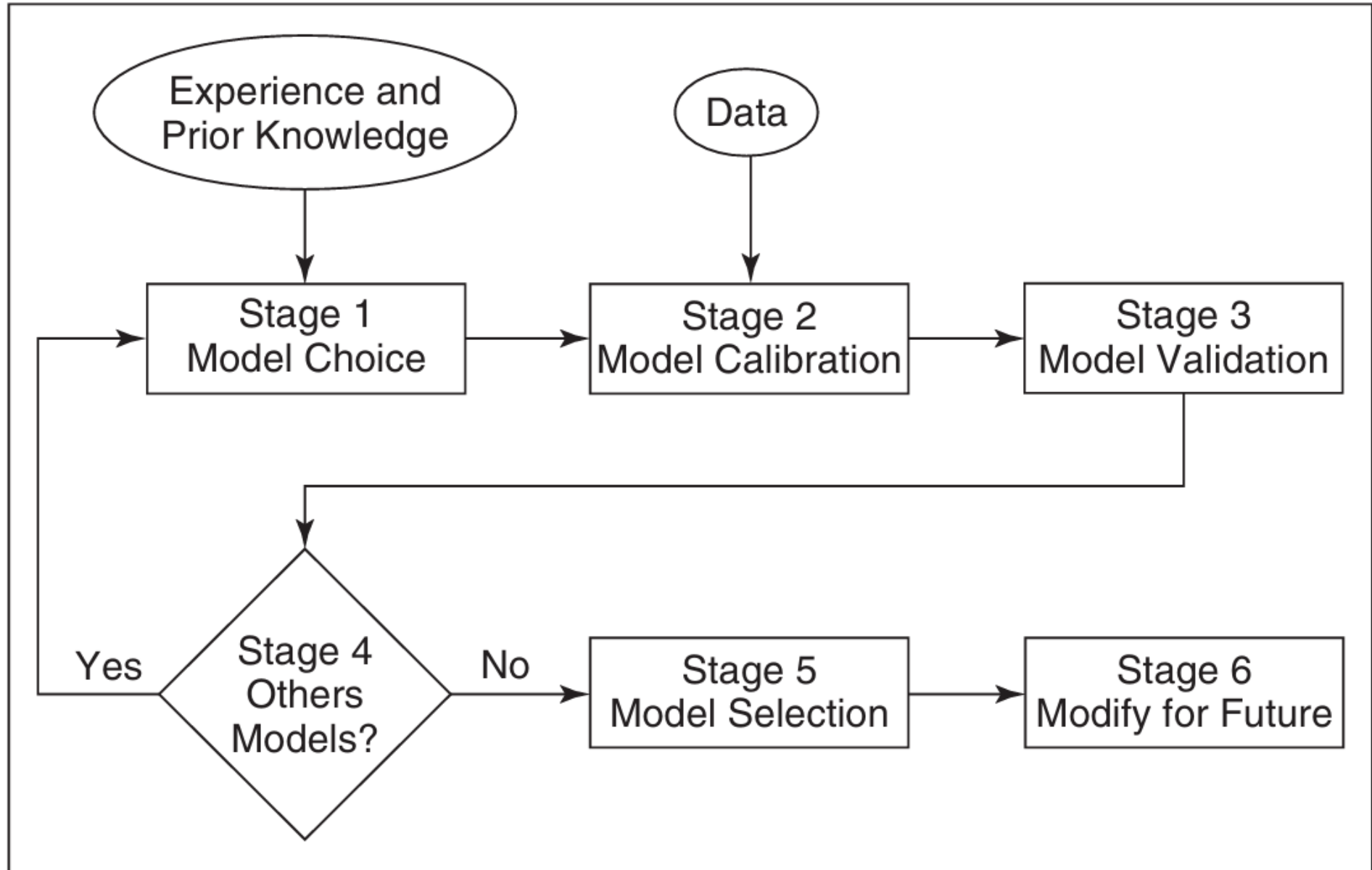
- ❑ Many problems in actuarial science involve the building of a **mathematical model** that can be used to **forecast or predict** insurance costs in the future.
- ❑ A model is mathematical description constructed based on the **knowledge** and **experience** of the actuary combined with **data from the past**.
- ❑ Nonlife insurance policies covering a fixed period of time, such as vehicle insurance, workers compensation insurance, and health insurance.
- ❑ The **two components** of claim losses,
  1. **claim frequency**: The number of claims in the block of policies is called the claim frequency
  2. **claim severity** or claim size: the monetary amount of each claim
- ❑ Both are modeled separately and combined in an aggregate loss distribution.
  - Claim frequency is modeled as a nonnegative discrete random variable,
  - Claim severity is continuously distributed.

## Learning objectives

- 1 Discrete distributions for modeling claim frequency
- 2 Binomial, geometric, negative binomial, and Poisson distributions
- 3  $(a,b,0)$  class of distributions
- 4 Compound distribution
- 5 Mixture distribution
- 6 Risk Models Revision

# The modeling process

□ Modeling process has six stage



# **Probability Distributions**

# Random Variables

---

❑ Before we can **handle the outcomes of a random process** like pricing an insurance policy, we must define this process using **three mathematical components**:

1. **Random Experiment**: A process with an **uncertain outcome**.
  - Actuarial Example: Observing a single policyholder for one year.
2. **Sample Space** ( $\Omega$ ): The set of **all possible outcomes** that could be observed.
  - Actuarial Example:  $\Omega = \{No\ Claim, Minor\ Accident, Total\ Loss, Theft, \dots\}$ .
3. **Event** ( $A$ ): A specific subset of the sample space we care about.
  - Actuarial Example: Event  $A = \{Any\ claim\ exceeding\ 10,000\}$

❑ The bridge between the three components is the **random variable**.

❑ In simple terms, a **random variable** is a **rule(mathematical function)** that maps each possible outcomes of an experiment into numbers that we can use for math and statistics.

$$X: \Omega \rightarrow \mathcal{R}$$

❑ In insurance, we usually split the random experiment into two distinct variables:

- Frequency ( $N$ )
  - How many times did the experiment result in a claim?
  - A discrete random variable.
- Severity ( $X$ ):
  - Given that a claim occurred, how much did it cost?"
  - A continuous random variable

# Random Variables

□ We can summarize the concept as follows

Random Experiment (A car drives for a year)

↓ Outcome (A crash occurs)

↓ Random Variable (The claim is 5,000)

↓ Distribution (The pattern of all such crashes)

↓ Actuarial Decision (Set the premium at 600).

| 1. The Random Experiment                             | 2. The Random Variable (X)                                                                                   | 3. The Probability Distribution                                                             |
|------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| The "Input" (Real World)                             | The "Bridge" (The Rule)                                                                                      | The "Output" (The Math)                                                                     |
| You observe a car for one year to see if it crashes. | Assign a number to the outcome: <b>0</b> for no crash, <b>1</b> for a minor dent, <b>2</b> for a total loss. | You assign a weight to those numbers: $P(X=0)=0.90$ , $P(X=1)=0.08$ , $P(X=2)=0.02$ .       |
| <b>Raw Reality:</b> The result is "a minor dent."    | <b>Translation:</b> The function $X(dent)$ outputs the number <b>1</b> .                                     | <b>Analysis:</b> Now you can calculate the <b>Expected Cost</b> using the number <b>1</b> . |

# Random Variables in Actuarial Science

---

□ Examples of random variables that might be encountered in actuarial work:

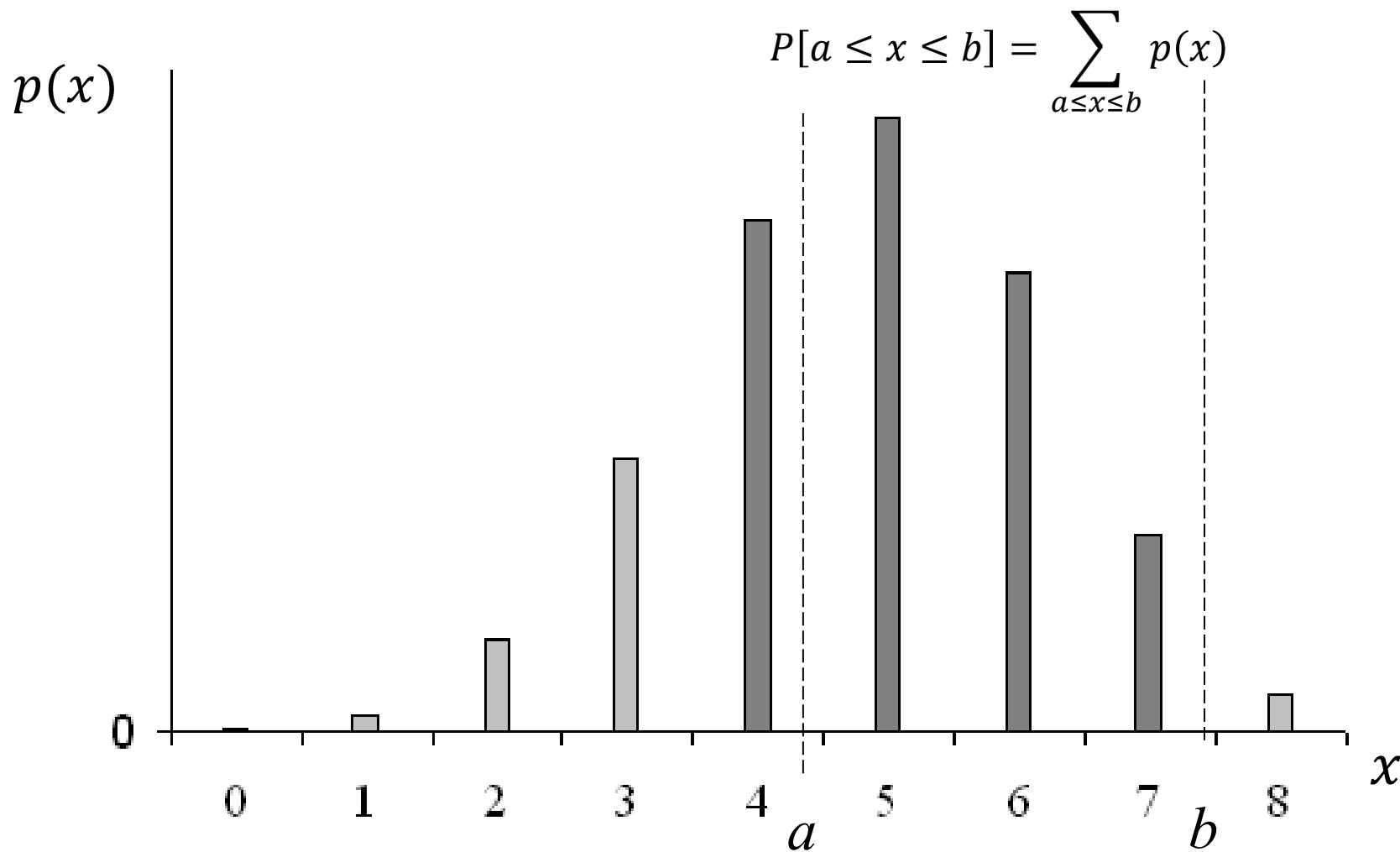
1. The age at death of a randomly selected birth,
2. The amount of money paid on a randomly selected automobile bodily injury claim,
3. The number of automobile bodily injury claims in one year from a randomly selected insured automobile,
4. The cost of medical malpractice claims paid in one year owing to events at a randomly selected hospital,

Because all of these phenomena can be expressed as random variables, the machinery of probability and mathematical statistics is at our disposal both to create and to analyze models for them

# Discrete Random Variables

□ For a discrete random variable  $X$  the probability distribution is described by the probability function  $p(x)$ , which has the following properties

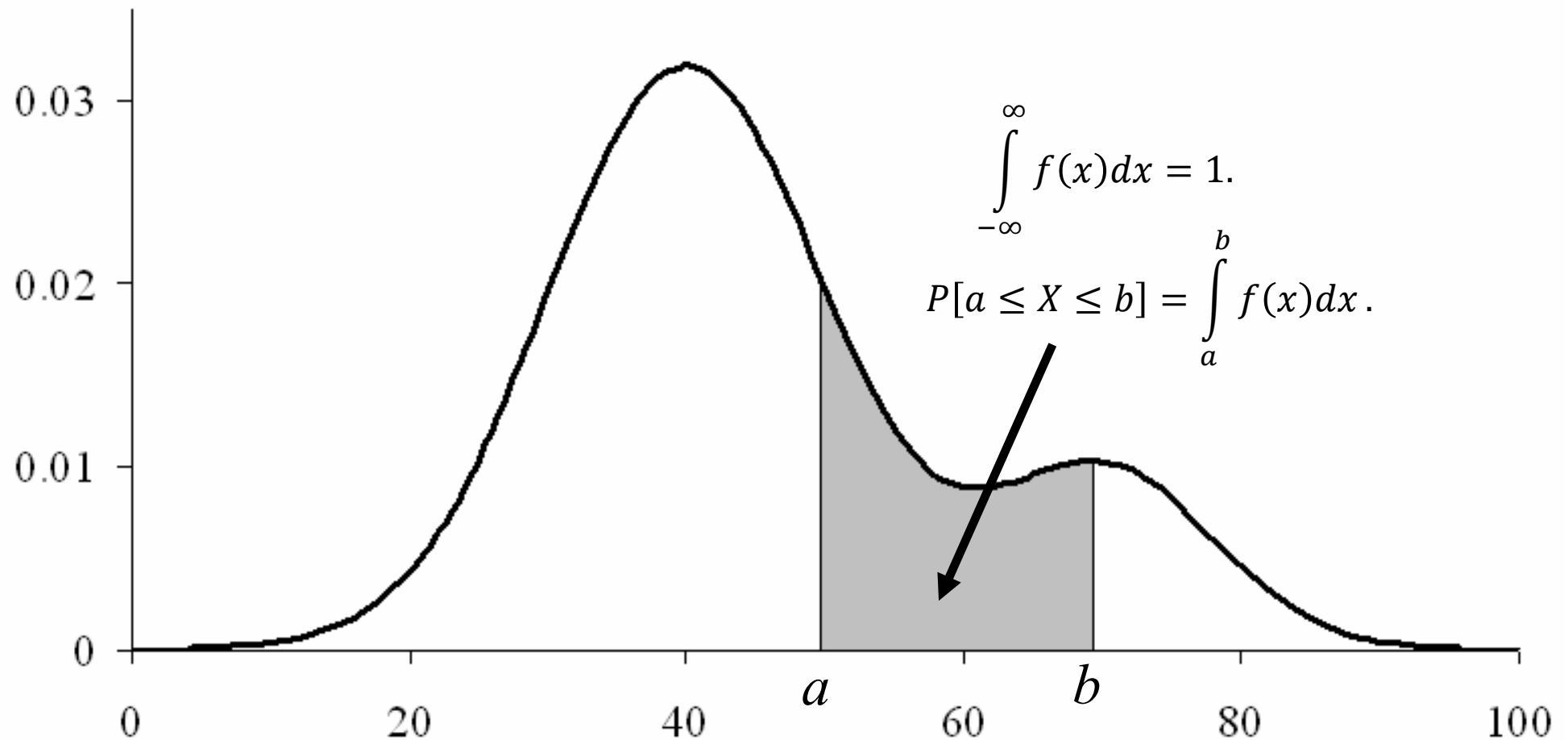
1.  $0 \leq p(x) \leq 1$
2.  $\sum_x p(x) = \sum_{i=1}^{\infty} p(x_i) = 1$
3.  $P[a \leq x \leq b] = \sum_{a \leq x \leq b} p(x)$



# Continuous random variables

□ For a continuous random variable  $X$  the probability distribution is described by the probability density function  $f(x)$ , which has the following properties :

1.  $f(x) \geq 0$ ,
2.  $\int_{-\infty}^{\infty} f(x)dx = 1$ ,
3.  $P[a \leq X \leq b] = \int_a^b f(x)dx$



# **Key Functions**

# Distribution Function

□ Distribution function (df) of random variable  $X$

$$F_X(x) = \Pr(X \leq x)$$

where  $\{X \leq x\}$  = the set of all outcomes (event) with  $X \leq x$

□ Probability density function (pdf) of continuous random variable

$$f_X(x) = \frac{dF_X(x)}{dx}$$

□ Probability mass function (pmf) of discrete random variable

$$f_X(x) = f(x) = \begin{cases} \Pr(X = x), & x \in \Omega \\ 0, & \text{otherwise} \end{cases}$$

## Properties

1.  $F(-\infty) = 0$  and  $F(\infty) = 1$ .
2.  $F(x)$  is non-decreasing (i. e. if  $x_1 < x_2$  then  $F(x_1) \leq F(x_2)$  )
3.  $F(b) - F(a) = P[a < X \leq b]$ .
4.  $p(x) = P[X = x] = F(x) - F(x-)$ ,  $F(x-) = \lim_{u \rightarrow x-} F(u)$
5. If  $p(x) = 0$  for all  $x$  (i.e.  $X$  is continuous) then  $F(x)$  is continuous.

# Distribution Function

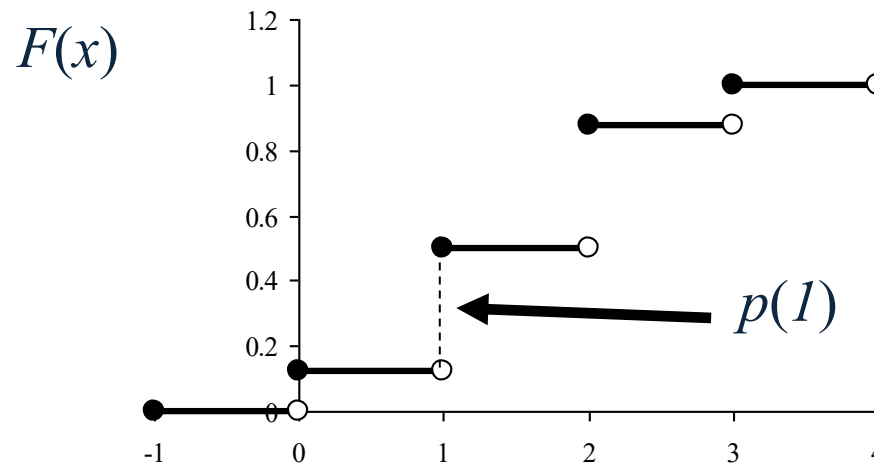
## 6. For Discrete Random Variables

$$F(x) = P[X \leq x] = \sum_{u \leq x} p(u)$$

$F(x)$  is a non-decreasing step function with

$$F(-\infty) = 0 \text{ and } F(\infty) = 1$$

$$p(x) = F(x) - F(x-) = \text{jump in } F(x) \text{ at } x.$$



# Distribution Function

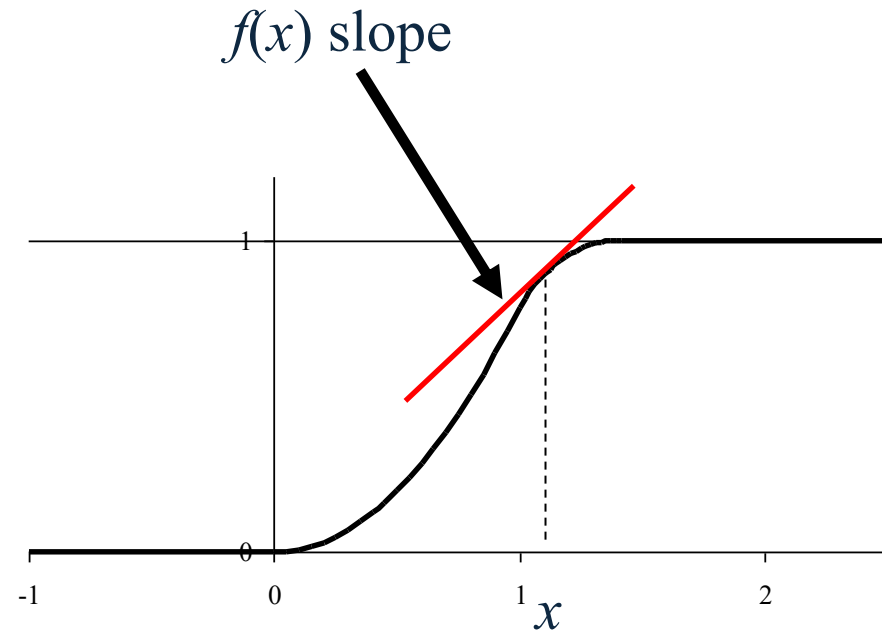
## 7. For Continuous Random Variables

$$F(x) = P[X \leq x] = \int_{-\infty}^x f(u) du$$

$F(x)$  is a non-decreasing continuous function with

$$F(-\infty) = 0 \text{ and } F(\infty) = 1$$

$$f(x) = F'(x).$$



To find the probability density function,  $f(x)$ , first find  $F(x)$  then  $f(x) = F'(x)$ .

# Survival Function

□ The survival function, usually denoted  $S_X(x)$ , for a random variable  $X$  is the probability that  $X$  is greater than a given number.

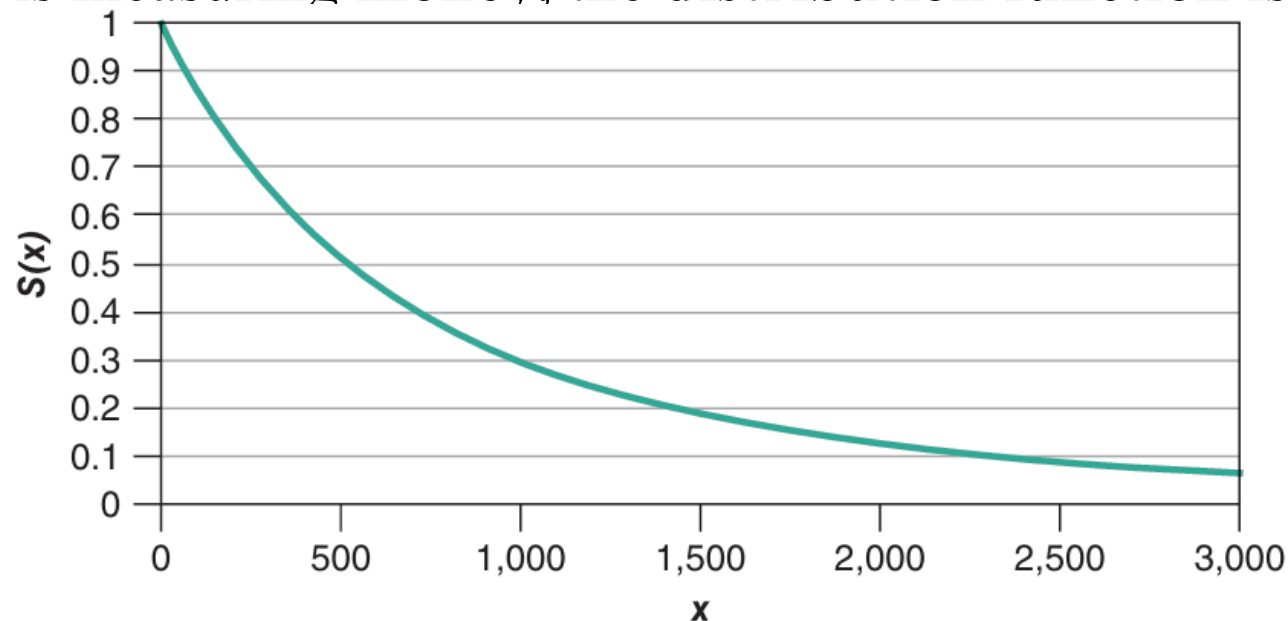
That is,

$$S_X(x) = P(X > x) = 1 - F_X(x).$$

## □ Properties

1.  $0 \leq S(x) \leq 1$  for all  $x$ ,
2.  $S(x)$  is non-increasing,
3.  $S(x)$  is right-continuous
4.  $\lim_{x \rightarrow -\infty} S(x) = 1$  and  $\lim_{x \rightarrow \infty} S(x) = 0$

□ When the random variable is measuring **time**, the **survival function** is presented, while when it is measuring **money**, the **distribution function** is presented.



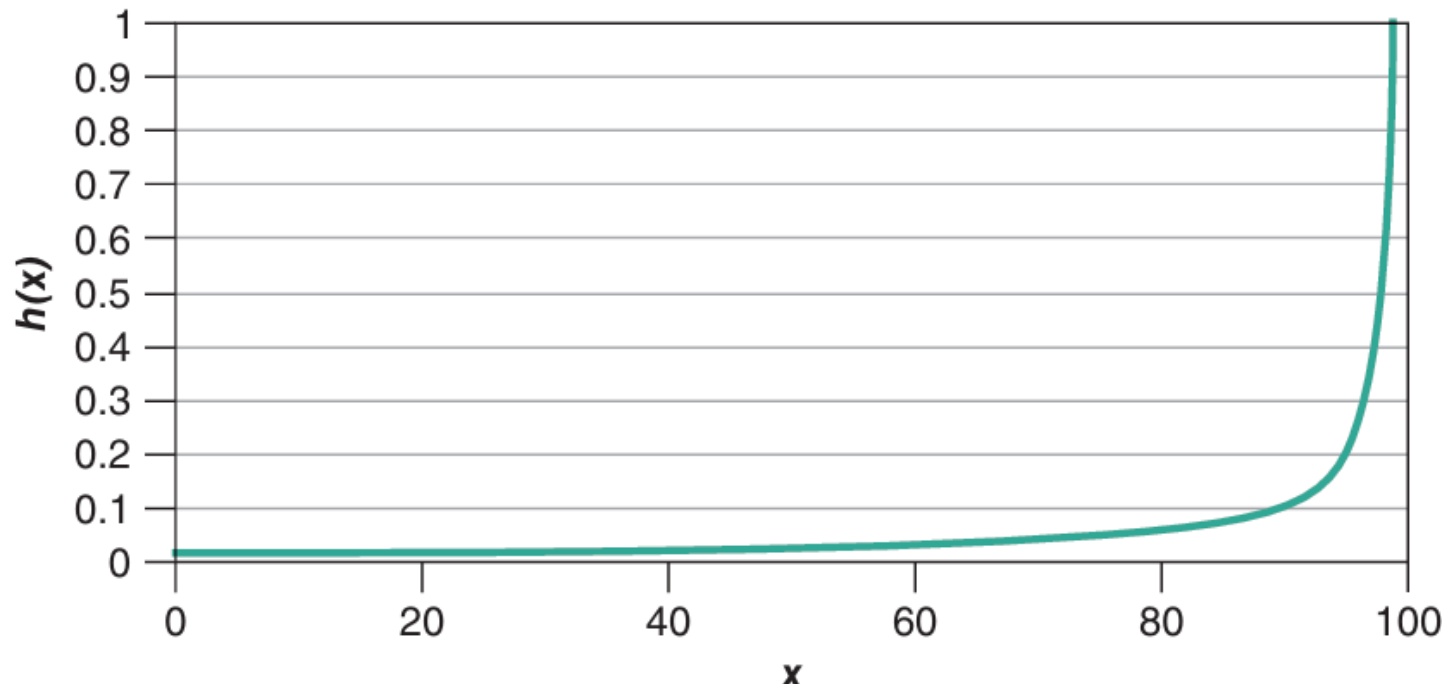
# Hazard Rate

- ❑ The hazard rate, also known as the **force of mortality** and the **failure rate** and usually denoted  $h_X(x)$ , is the ratio of the density and survival functions when the density function is defined.

That is,

$$h_X(x) = \frac{f_X(x)}{S_X(x)}.$$

- ❑ It may be interpreted as the probability density at  $x$  given that the argument will be at least  $x$ .
- ❑ In mortality terms, the force of mortality is the annualized probability that a person age  $x$  will die in the next instant, expressed as a death rate per year,



# **Basic Distributional Quantities**

# Expectation

---

□ For a discrete random variable  $X$  with probability mass function  $p(x)$ , the expected value of  $X$  denoted as  $E(X)$  or  $\mu$  is defined to be:

$$E(X) = \sum_x xp(x) = \sum_i x_i p(x_i)$$

□ If  $X$  is **continuous** with probability density function  $f(x)$ , then

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

□ The expected value of a function  $g(X)$ ,  $E[g(X)]$  is defined to be:

$$E[g(X)] = \sum_x g(x)p(x)$$

□ If  $X$  is **continuous** with probability density function  $f(x)$

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

# Variance

---

- ❑ The variance of a random variable  $X$ , denoted as  $Var(X)$  or  $\sigma^2$ , is defined as the **expected value of the squared deviation from the mean.**

$$Var(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2$$

- ❑ For discrete random variable

$$Var(X) = \sum_i (x_i - \mu)^2 P(X = x_i)$$

- ❑ For Continuous Variables

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

- ❑ If the expected value tells you where the center of the function's output is, the variance tells you how much the actual outcomes deviate from that center on average.

# Moments of a Random Variable

---

- ❑ **Moments** describe the "shape" of the probability mass.
- ❑ It is the expected value of a power of  $X$  about a given point
- ❑ Types of moments

## 1. Raw Moments (About the Origin)

$$\mu'_k = E(X^k) = \begin{cases} \sum_x x^k p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x^k f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

- The first moment at  $k = 1$ ,  $\mu_1 = E(X)$  is the center of gravity of the distribution of  $X$ .

## 2. Central Moments (About the Mean)

$$\mu_k = E((X - \mu)^k) = \begin{cases} \sum_x (x - \mu)^k p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

- ❑ The 2<sup>nd</sup> Central Moment is the Variance ( $\sigma^2$ ).
- ❑ The ratio of the standard deviation to the mean is called the coefficient of variation.
- ❑ The ratio of the third central moment to the cube of the standard deviation is called the **skewness**.
- ❑ The ratio of the fourth central moment to the fourth power of the standard deviation is called the **kurtosis**.

# Moment generating functions

---

- ❑ A **Moment Generating Function (MGF)** is a powerful mathematical tool that essentially collects all the information about a random variable's distribution into a single function.
- ❑ Let  $X$  denote a random variable, Then the moment generating function of  $X$ ,  $m_X(t)$  is defined by:

$$m_X(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

# Properties of Moment Generating Functions

---

1.  $m_X(0) = 1$
2.  $m_X(t) = 1 + \mu_1 t + \frac{\mu_2}{2!} t^2 + \frac{\mu_3}{3!} t^3 + \dots + \frac{\mu_k}{k!} t^k + \dots$ 
  - Taylor expansion of  $e^{tX}$  is  $1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \dots$ .
3.  $m_X^{(k)}(0) = \mu_k = E(X^k)$   $k^{th}$  derivative of  $m_X(t)$  at  $t = 0$ .
4. If  $Y$  is a random variable such that  $Y = bX + a$ , then
$$m_Y(t) = E(e^{(bX+a)t}) = e^{at} E(e^{bXt}) = e^{at} m_X(bt)$$
5. If  $X$  and  $Y$  are two independent random variables with moment generating function  $m_X(t)$  and  $m_Y(t)$ , then
$$\begin{aligned} m_{X+Y}(t) &= E(e^{[X+Y]t}) = E(e^{Xt} e^{Yt}) \\ &= E(e^{Xt}) E(e^{Yt}) \\ &= m_X(t) m_Y(t) \end{aligned}$$

# **Some Important Discrete distributions**

Model for Actuarial

# The Bernoulli distribution

---

□ Suppose that we have an experiment that has two outcomes  
Success (S) and Failure (F)

□ These terms are used in reliability testing, Suppose that  
 $p$  is the probability of success (S) and  
 $q = 1 - p$  is the probability of failure (F)

This experiment is sometimes called a **Bernoulli Trial**

If  $X = \begin{cases} 0 & \text{if the outcome is F} \\ 1 & \text{if the outcome is S} \end{cases}$  then  $p(x) = P[X = x] = \begin{cases} q & x = 0 \\ p & x = 1 \end{cases}$

□ For an actuary,

□ A single life insurance policy for one year can be modeled as a Bernoulli trial.

- Success (1): A claim is paid.
- Failure (0): No claim is paid.

□ Policy Lapses:

- Whether a specific customer chooses to renew their policy or let it "lapse" at the end of the term.

□ Loan Default:

- In credit risk, whether a single borrower defaults on their mortgage this month (1) or makes their payment (0).

# The Binomial distribution

---

- We observe a Bernoulli trial (**S**, **F**)  $n$  times.
- Let  $X$  denote the number of successes in the  $n$  trials. Then  $X$  has a binomial distribution,

$$p(x) = P[X = x] = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, \dots, n$$

$$E(X) = np \text{ and } \text{Var}(X) = npq$$

Where:

1.  $p$  = the probability of success (**S**), and
  2.  $q = 1 - p$  = the probability of failure (**F**)
- An insurance company has a small portfolio of 100 identical homes ( $n = 100$ ) in a flood-prone area. Based on historical data, the probability of any single home filing a flood claim in a year is 2% ( $p = 0.02$ ). The actuary uses the Binomial distribution to answer:

Expected Number of Claims:  $E[X] = np = 100 \times 0.02 = 2$  claims.

The "Tail Risk": What is the probability that more than 5 homes flood?

$$P(X > 5) = 1 - P(X \leq 5)$$

- This calculation is vital for determining if the company has enough reserves (cash on hand) to pay out in a worst-case scenario.

# The Poisson Distributions

---

- ❑ Suppose events are occurring randomly and uniformly in time.
- ❑  $X$  = the number of events occurring in a **fixed period of time**.
- ❑  $X \sim$  **Poisson** distribution with parameter  $\lambda$  with probability function as:

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x = 0, 1, 2, 3, 4, \dots$$

$$E(X) = \lambda \text{ and } \text{Var}(X) = \lambda$$

- ❑ The Poisson distribution is the most common model for claim frequency in a fixed interval of time or space.
- ❑ Unlike the Binomial, it does not consider number of policies; you simply count how many events occur within fixed interval of time.
- ❑ **Death Claims:**
  - A small life insurance company might receive an average of 6 death claims per day ( $\lambda = 6$ ).
  - The Poisson model calculates the probability of receiving a spike in claims (e.g., 10 or more) on a single day.
- ❑ **Accident Analysis:**
  - Modeling the number of traffic accidents on a specific road during a month.
- ❑ **Operational Risk:**
  - Estimating the number of "rare events" like machine malfunctions or fraudulent transactions in a year.

# The Geometric Distributions

---

□ Suppose a Bernoulli trial (S,F) is repeated **until a success occurs**.

□  $X$  = the trial on which the **first success (S) occurs**.

□  $X \sim$  **geometric** distribution with probability function of  $X$  as:

$$p(x) = P[X = x] = (1 - p)^{x-1}p = p q^{x-1}$$

$$E(X) = \frac{q}{p} \text{ and } \text{Var}(X) = \frac{q}{p^2}$$

□ The Geometric distribution models the "waiting time" until the first success (or claim) occurs in a series of independent trials.

□ **Policy Lapse Timing:**

- An actuary might model how many years a policy holder remains with a company before their first "lapse" (canceling the policy).

□ **Equipment Failure:**

- In industrial insurance, modeling the number of times a safety switch is operated before it fails for the first time.

□ **Reliability Testing:**

- The number of server requests handled before the first system failure occurs.

□ The **only** discrete distribution with the **Memoryless Property**.

$$P(X > s + t \mid X > s) = P(X > t)$$

The "age" of the policy doesn't change the future risk.

# The Negative Binomial Distributions

---

□ Suppose a Bernoulli trial (S,F) is repeated **until  $k$  successes** occur.

□  $X$  = the trial on which the  $k^{\text{th}}$  success (S) occurs.

□  $X \sim$  **negative binomial** with probability function as:

$$p(x) = P[X = x] = \binom{x-1}{k-1} p^k q^{x-k} \quad x = k, k+1, k+2, \dots$$

$$E(X) = \frac{kq}{p} \text{ and } \text{Var}(X) = \frac{kq}{p^2}$$

□ The Negative Binomial is often used as a "superior" version of the Poisson. While Poisson assumes the mean equals the variance, real-world insurance data often shows **overdispersion** ( $\frac{kq}{p^2} > \frac{kq}{p}$ ).

□ **Accident Proneness:**

- Actuaries use this to account for the fact that some individuals have a higher "propensity" for accidents than others. It models the total number of failures before a set number of "successes" (accident-free months) is reached.

□ **Motor Insurance Ratemaking:**

- Modeling claim counts when the population is heterogeneous (diverse). For example, if some drivers are significantly riskier than others, a Poisson model will fail, but a Negative Binomial will accurately capture the "fat tail" of high claim counts.

□ **Bonus-Malus Systems:**

- Used in calculating premium discounts or surcharges based on the number of claims a driver has filed.

# **Continuous Distributions**

# The Uniform distribution from a to b

---

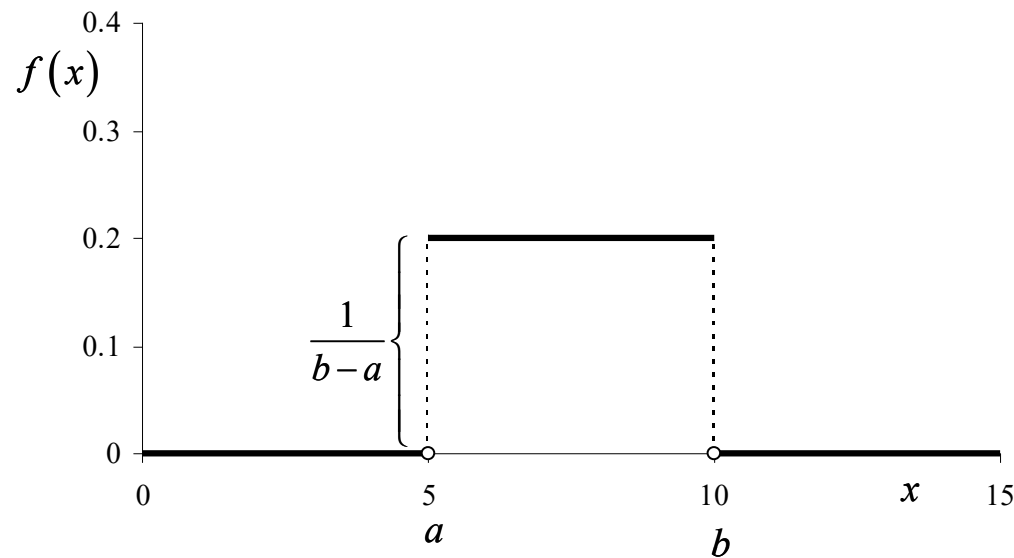
□ A random variable  $X$  follows a continuous uniform distribution on the interval  $[a, b]$ , denoted as  $X \sim U(a, b)$ , if its Probability Density Function (PDF) is constant:

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

$$E[X] = \frac{a+b}{2} \text{ and } Var(X) = \frac{(b-a)^2}{12}$$

□ All claim amounts between 1,000 and 5,000 are equally likely

□ The probability that the claim lies in any interval depends only on the width of the interval



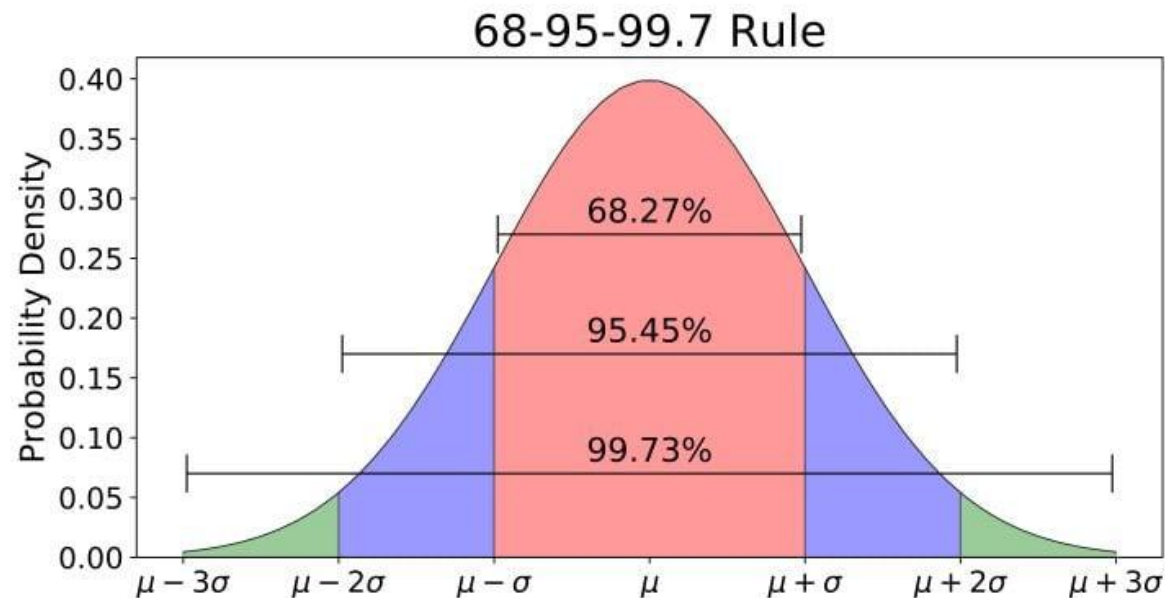
# The Normal distribution

□ A continuous random variable  $X$  is said to follow a **Normal distribution** with mean  $\mu$  and variance  $\sigma^2$  if its probability density function is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

□ In actuarial science, the normal distribution is used to model **aggregate quantities** that result from the **sum of many small, independent risks**, such as:

- Total claims over a large portfolio
- Aggregate losses over a period
- Average claim size in large samples
- Approximation of other distributions (via CLT)

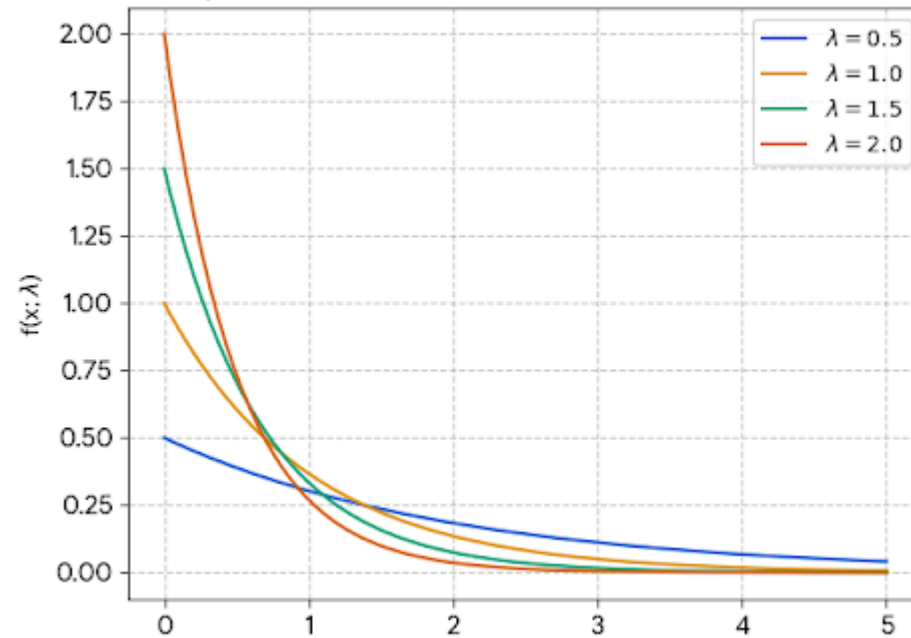


# The Exponential distribution

- A continuous random variable  $X$  is said to follow an **Exponential distribution** with **rate parameter**  $\lambda > 0$  if its probability density function is:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$
$$E[X] = \frac{1}{\lambda} \text{ and } Var(X) = \frac{1}{\lambda^2}$$

- It is used for modeling the **time between events**.
- It is used in modeling the **frequency** and **severity** of risks.
  - If claims arrive at a **constant rate**  $\lambda$ , waiting time until next claim has an  $Exp(\lambda)$  distribution.
  - used to model claim sizes because
- The future waiting time does **not depend on the past** ”Memoryless Property”



# Weibull distribution

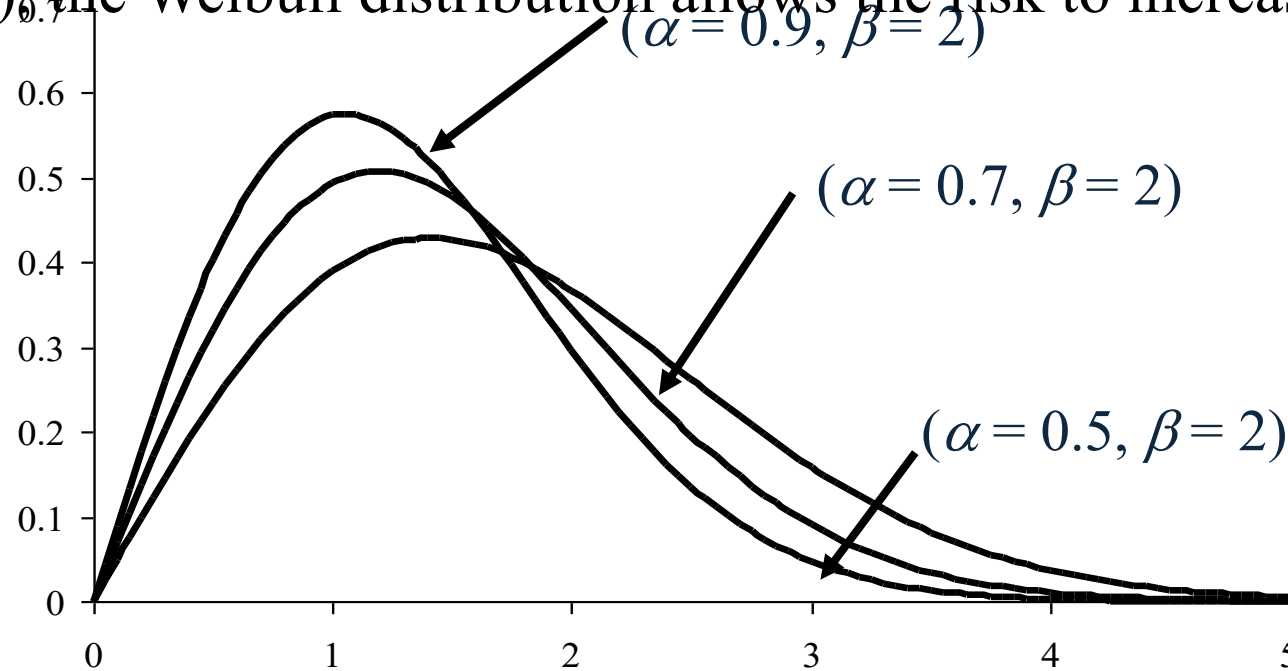
□ **Weibull distribution** is a generalization of the exponential distribution, parameterized with  $\alpha$  (shape) and  $\beta$  (scale).

$$f(x) = \frac{\alpha}{\beta} \left( \frac{x}{\beta} \right)^{\alpha-1} e^{-(x/\beta)^\alpha}, x \geq 0$$

- $\alpha$ : "aging" parameter,
- $\beta$ : characteristic life, higher  $\beta$  indicates larger expected claim sizes

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

□ While the exponential distribution assumes a constant "hazard rate" (risk doesn't change over time), the Weibull distribution allows the risk to increase or decrease.



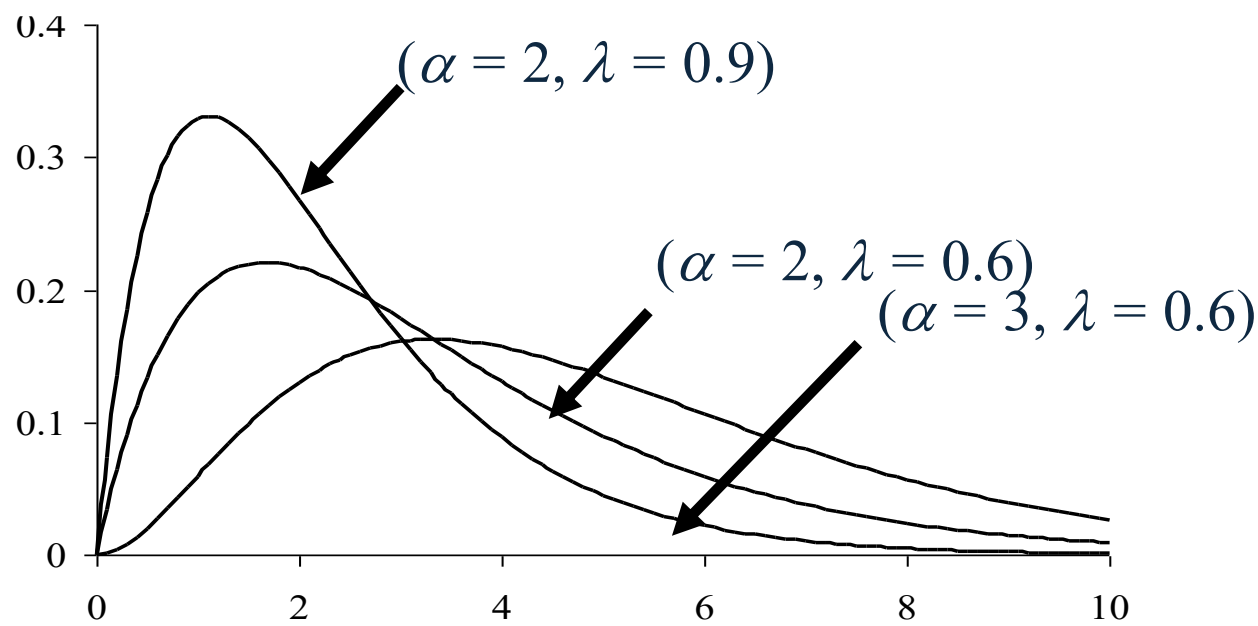
# The Gamma distribution

Let the continuous random variable  $X$  have density function:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases},$$

Then  $X$  is said to have a **Gamma distribution** with parameters  $\alpha$  and  $\lambda$ .

- If the time between individual claims follows an Exponential distribution with rate  $\lambda$ , then the time until the  $\alpha^{\text{th}}$  claim occurs follows a  $\text{Gamma}(\alpha, \lambda)$  distribution
- Gamma distribution is most commonly used to model total claim amounts and claim frequency over a specific time period



# The Gamma distribution

---

1. The set of gamma distributions is a *family* of distributions (parameterized by  $\alpha$  and  $\lambda$ ).
2. Contained within this family are other distributions
  - a. **The Exponential distribution** – in this case  $\alpha = 1$ , the gamma distribution becomes the exponential distribution with parameter  $\lambda$ . The exponential distribution arises if we are measuring the lifetime,  $X$ , of an object that does not age. It is also used a distribution for *waiting times* between *events* occurring *uniformly* in time.
  - b. **The Chi-square distribution** – in the case  $\alpha = \nu/2$  and  $\lambda = 1/2$ , the gamma distribution becomes the chi- square ( $\chi^2$ ) distribution with  $\nu$  degrees of freedom. Later we will see that sum of squares of independent standard normal variates have a chi-square distribution, degrees of freedom = the number of independent terms in the sum of squares.

# The (a, b, 0) Class of Distributions

- ❑ It consists of discrete probability distributions where the **ratio of successive probabilities** can be expressed as a **linear function** of the number of events  $k$ .
- ❑ A nonnegative discrete random variable  $X$  is in the  $(a, b, 0)$  class if its probability mass function  $p_x = f_X(x)$  satisfies the following recursion

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}$$

- ❑ where  $a$  and  $b$  are constants, that define the specific distribution within the class.
- ❑ The "0" in  $(a, b, 0)$  indicates that the recursion starts at  $k = 1$ , meaning the probability at zero ( $p_0$ ) is defined by the same formula as the rest of the distribution.
- ❑ Actuaries choose between the following distribution

| Distribution      | a                       | b                            | Actuarial Use Case                                            |
|-------------------|-------------------------|------------------------------|---------------------------------------------------------------|
| Poisson           | 0                       | $\lambda$                    | Standard frequency for independent events.                    |
| Binomial          | $\frac{-q}{1-q}$        | $\frac{(m+1)q}{1-q}$         | When the number of claims has a hard upper limit.             |
| Negative Binomial | $\frac{\beta}{1+\beta}$ | $(r-1)\frac{\beta}{1+\beta}$ | <b>Over-dispersed</b> data (e.g., heterogeneous risk groups). |

# Some Methods for Creating New Distributions

---

## 1. Compound distribution

□ Let  $X_1, \dots, X_N$  be iid nonnegative integer-valued random variables, each distributed like  $X$ . We denote the sum of these random variables by  $S$ , so that

$$S = X_1 + \dots + X_N.$$

□ If  $N$  is itself a nonnegative integer-valued random variable distributed independently of  $X_1, \dots, X_N$ , then  $S$  is said to have a **compound distribution**.

□ The distribution of  $N$  is called the **primary distribution**, and the distribution of  $X$  is called the **secondary distribution**.

□ If  $N$  is Poisson and  $X$  is geometric,  $S$  has a Poisson-geometric distribution.

□ A compound Poisson distribution is a compound distribution where  $N$  is Poisson, for any secondary distribution.

# Some Methods for Creating New Distributions

## 2. Mixture distribution

□ Let  $X_1, \dots, X_n$  be random variables with corresponding pf or pdf  $f_{X_1}(\cdot), \dots, f_{X_n}(\cdot)$  in the common support  $\Omega$ .

□ A new random variable  $X$  may be created with pf or pdf  $f_X(\cdot)$  given by

$$f_X(x) = p_1 f_{X_1}(x) + \dots + p_n f_{X_n}(x), \quad x \in \Omega,$$

where  $p_i \geq 0 \forall i = 0, 1, \dots, n$  and  $\sum_{i=0}^n p_i = 1$

□ The mean of  $X$  is

$$E(X) = \mu = \sum_{i=1}^n p_i \mu_i$$

While

$$Var(X) = \sum_{i=1}^n p_i [(\mu_i - \mu)^2 - \sigma^2]$$

# **Risk Models**

# Risk Models

---

- ❑ Risk models describe the risk in a form of a random variable, so that we can figure out the **likelihood** of a loss and **how it can occur**.
- ❑ But, we need also to quantify how much a person's **aversion for risk**.
- ❑ An actuary uses **utility function** to explain **why people buy insurance** as trade-offs between **premium paid** and **financial security**.
- ❑ risk-appetite has three types based on the shape of the utility function  $u(w)$ :
  - Risk-Averse (Concave):  $u''(w) < 0$ .
  - Risk-Neutral (Linear):  $u''(w) = 0$ .
  - Risk-Seeking (Convex):  $u''(w) > 0$ .
- ❑ A risk-averse person is willing to pay a premium ( $P$ ) that is greater than the expected loss ( $E[L]$ ).
  - Pure Premium:  $E[L]$
  - Risk Premium: The "extra" amount paid to avoid the volatility of the loss.
- ❑ The utility of your "average" wealth  $u(E[w])$  is higher than the average utility of your "volatile" wealth  $E[u(w)]$ .
$$u(E[w]) > E[u(w)] \text{ (Jensen's Inequality)}$$
- ❑ This mathematical gap is where the insurance industry lives and breathes.

# Risk Models

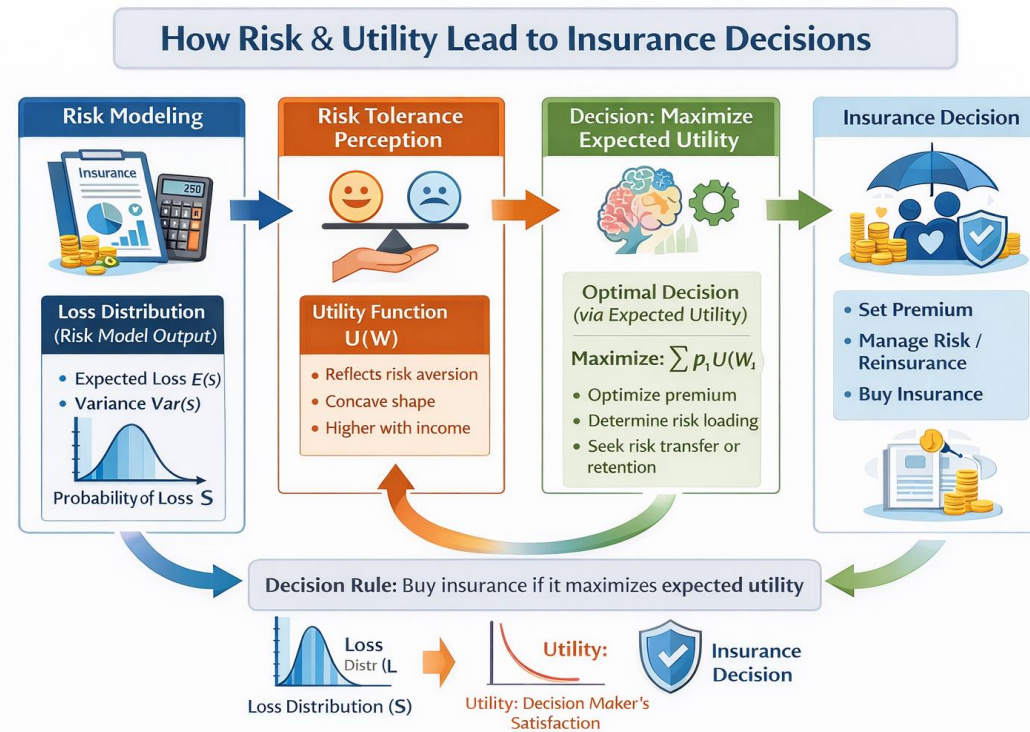
❑ Different mathematical functions model different economic behaviors

| Function Type | Formula                        | Characteristic                               |
|---------------|--------------------------------|----------------------------------------------|
| Linear        | $u(w) = aw + b$                | Risk-neutral; used for large corporations.   |
| Exponential   | $u(w) = \frac{1 - e^{-aw}}{a}$ | Constant Absolute Risk Aversion (CARA).      |
| Logarithmic   | $u(w) = \ln(w)$                | Risk aversion decreases as wealth increases. |
| Power         | $u(w) = w^\gamma$              | Used in complex financial modeling.          |

❑ Utility theory is the primary tool used to decide how much risk a primary insurer should keep versus how much they should pass to a **reinsurer**.

| Component         | Role in the Relationship                                                              |
|-------------------|---------------------------------------------------------------------------------------|
| Risk Model        | Defines the "thickness" of the tail (how bad the worst-case scenario is).             |
| Utility Function  | Defines the "pain threshold" of the insurer's capital base.                           |
| Optimal Retention | The point where cost of reinsurance equals the utility gained from avoiding the risk. |

# Risk Models



❑ In Actuarial Science, we generally look at risk through two different lenses depending on whether we care about

1. The **individual policyholder**: Individual Risk Model
2. The **entire portfolio**: Collective Risk Model

❑ Purpose of Risk Models

- Price premiums
- Set reserves
- Measure risk
- Support credibility theory

# The Individual Risk Model

---

- ❑ In this model, we view the total loss as the sum of a fixed number of insurance units (e.g.,  $n$  independent policies).
- ❑ The Setup:
  - Each policy  $i$  has a loss  $X_i$ .
  - $X_i = I_i \times Y_i$
  - $I_i$  is an indicator:  $I_i = 1$  if policy  $i$  has a claim,  $I_i = 0$  otherwise
  - $Y_i$  is the severity of the claim
  - The **Total Loss**:  $S = X_1 + X_2 + \dots + X_n$
- ❑ Key Characteristic:
  - We assume  $n$  is known and fixed.
  - Each insured has **at most one claim (standard case)**
  - The "risk" comes from whether a claim occurs for each specific person and how much that claim is.
- ❑ Distribution:
  - Usually handled via the Bernoulli or Binomial distribution for the "occurrence" part.
$$E[S] = \sum_{i=1}^n E[X_i] \text{ and } Var(S) = \sum_{i=1}^n Var(X_i)$$
- ❑ Crucial Note:
  - This model is computationally "heavy" for large portfolios because you have to track every single policy separately.

# The Collective Risk Model

❑ The CRM is more complex because it uses a **Compound Distribution**.

The Setup:

- We decouple the frequency (how many) from the severity (how much).
- The Total Loss:  $S = \sum_{i=1}^N X_i$
- $N$ : random variable representing the number of claims (Frequency).
- $X_i$ : Random variables representing the size of each claim (Severity).

❑ Key Assumption:

- $N, X_1, X_2, \dots$  are Independent, and Identically Distributed (IID.).

$$S = X_1 + X_2 + \dots + X_N$$

❑ Distribution

| Frequency (N)     | Severity (X) | Resulting Compound Distribution (S) |
|-------------------|--------------|-------------------------------------|
| Poisson           | Gamma        | Compound Poisson                    |
| Poisson           | Exponential  | Compound Poisson                    |
| Negative Binomial | Logarithmic  | Negative Binomial                   |

❑ Using the law of total expectation and variance, we get the fundamental "Compound" formulas:

$$E[S] = E[N] \cdot E[X] \text{ and } Var(S) = E[N]Var(X) + Var(N)E[X]^2$$

# Panjer Recursion

---

- For the recursion to work, the frequency distribution (N) must satisfy the following relation:

$$p_k = P(N = k) = \left(a + \frac{b}{k}\right) \cdot p_{k-1}, \quad k = 1, 2, 3, \dots$$

- This is known as the  $(a, b, 0)$  class, which includes:

Poisson:  $a = 0, b = \lambda$

Binomial:  $a = \frac{-q}{1-q}, b = \frac{(n+1)q}{1-q}$

Negative Binomial:  $a = \beta/(1 + \beta), b = (r - 1)a$

- If the claim sizes (X) are discrete (or discretized) on the integers  $0, 1, 2, \dots$  with probability mass function  $f_x = P(X = x)$ , the probability of the total aggregate loss  $S = s$  is:

$$g_s = P(S = s) = \frac{1}{1 - af_0} \sum_{j=1}^s \left(a + \frac{bf_j}{s}\right) f_j \cdot g_{s-j}.$$

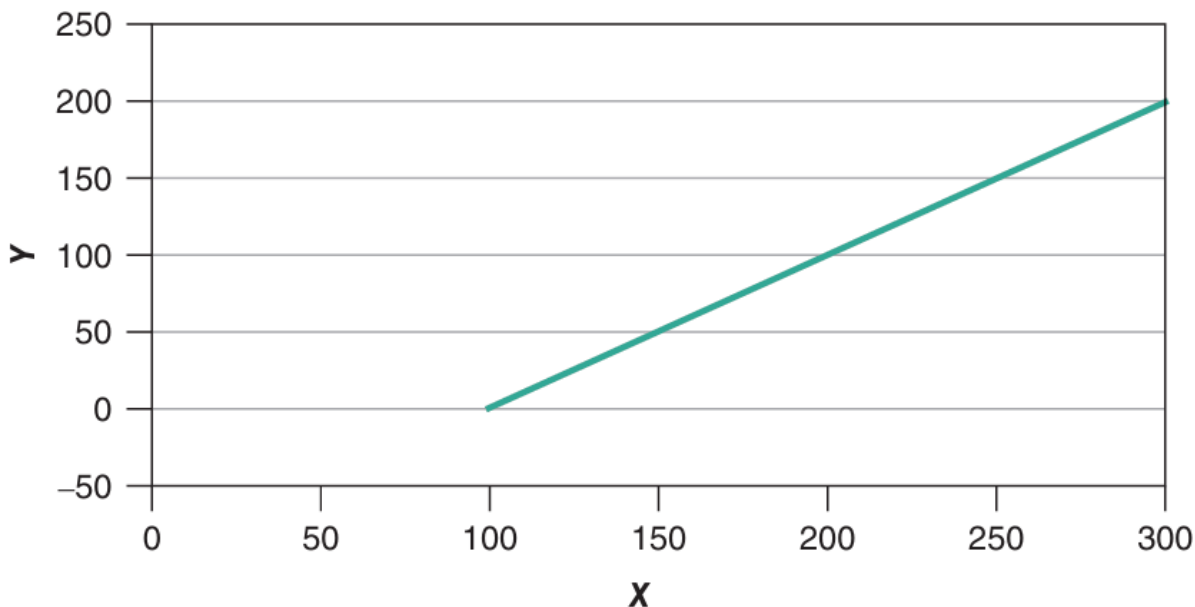
# Claim Modification

---

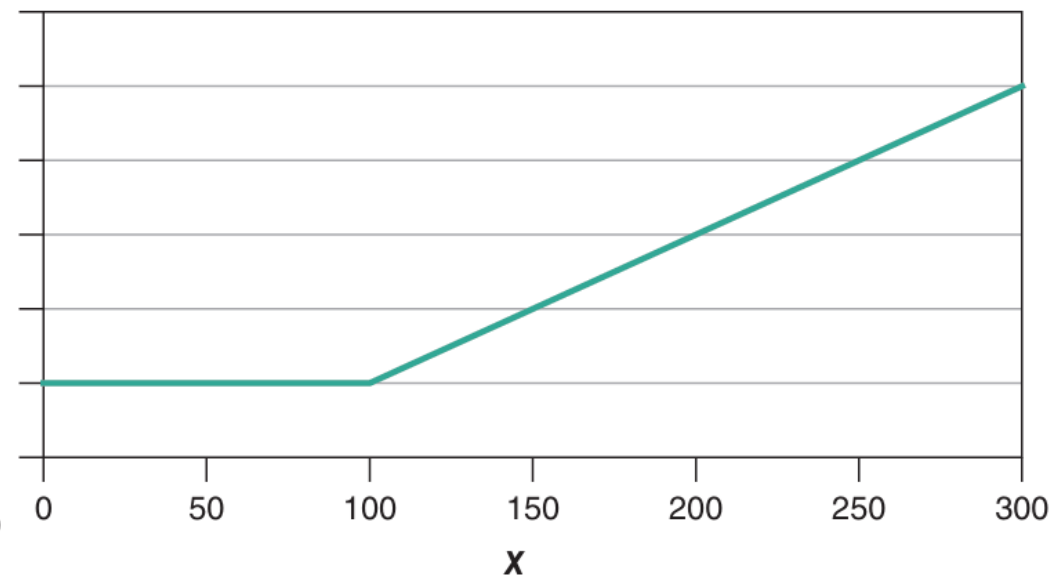
- ❑ There are two different ways to view the random variable of insurance payments:
  1. Payment per Payment Variable ( $Y^L$ ): Excess Loss Variable.
    - Only events that result in an **actual payment are recorded**; all others are **ignored/deleted** from the data set,
    - the variable exists only when a payment is made
  2. Payment per Loss Variable ( $Y^P$ ): Left-Censored and Shifted.
    - Every insured event (loss) is recorded, even if the payment is 0m
    - considers the value zero whenever a loss produces no payment
- ❑ Coverage Modifications
  - Explains how policy limits, and deductibles change the original loss distribution ( $X$ ).
- ❑ The Effects of Deductibles
  - Specifically focusing on how a threshold  $d$  transforms the mean and variance of the claims.
- ❑ The purpose of the **Deductible** is twofold:
  1. **Reducing Small Claims:** To prevent the company from opening a formal claim file for trivial amounts (like 100), where the administrative cost of processing the claim would exceed the value of the repair itself.
  2. **Encouraging Caution:** When you know you are responsible for the first 750 out of your own pocket, you are likely to be more careful to avoid minor accidents.

# Claim Modification

- ❑ Graphs illustrating the difference between the excess loss variable and the left censored and shifted variable.
- ❑ Excess loss: helps an insurer understand the severity of actual accidents. It answers the question: *"When I have to write a check, how big is it usually?"*
- ❑ Left censored and shifted: used to determine the total expected cost of insuring a group. It answers the question: *"On average, how much will I pay out for every policy I sell?"*



Excess loss variable



Left censored and shifted variable

# Excess Loss Variable

---

□ For a given value of  $d$  with  $P(X > d) > 0$ , the excess loss variable is

$$Y^P = X - d \text{ given that } X > d.$$

- $X - d$ : This is the "excess loss" (the amount the insurer pays after the deductible).
- $|mid X > d$ : condition that a payment is actually triggered because the loss exceeded the deductible.

□ This variable could also be called a **left truncated** (any values of  $X$  below  $d$  are not observed.) and **shifted variable** ( is subtracted from the remaining values.)

□ Its expected value,

$$e(d) = E(Y^P) = E(X - d | X > d),$$

is called the mean excess loss function.

□ Its variance

$$Var(Y^P) = \frac{E[(X - d)^2 | X > d] - [e(d)]^2}{S(d)} = \frac{\int_d^\infty (x - d)^2 f(x) dx}{S(d)} - [e(d)]^2$$

□ It is a crucial metric used to understand the severity of "large" claims.

□ If we already know a loss has surpassed the deductible  $d$ , how much more, on average, will the insurer have to pay?

# Left Censored and Shifted Variable

---

- ❑ The left-censored and shifted variable describes the **actual amount paid** across *all* possible losses.
- ❑ Let  $X$  be the **total loss**, if there is a deductible  $d$ , the insurer's payment variable, often denoted as  $Y^L$  (payment per loss), is defined as:

$$Y^L = (X - d)_+ = \max(0, X - d)$$

- Left-Censored: Any loss  $X$  that is less than or equal to  $d$  is "censored" not ignored but are set equal to zero.
- Shifted: For any loss  $X > d$ , the value is "shifted" downward by  $d$ . You aren't paying the whole loss; you are only paying the part that exceeds the threshold.

- ❑ Its expected value

$$E[(X - d)_+] = \int_d^{\infty} S(x) dx$$

- ❑ Its variance

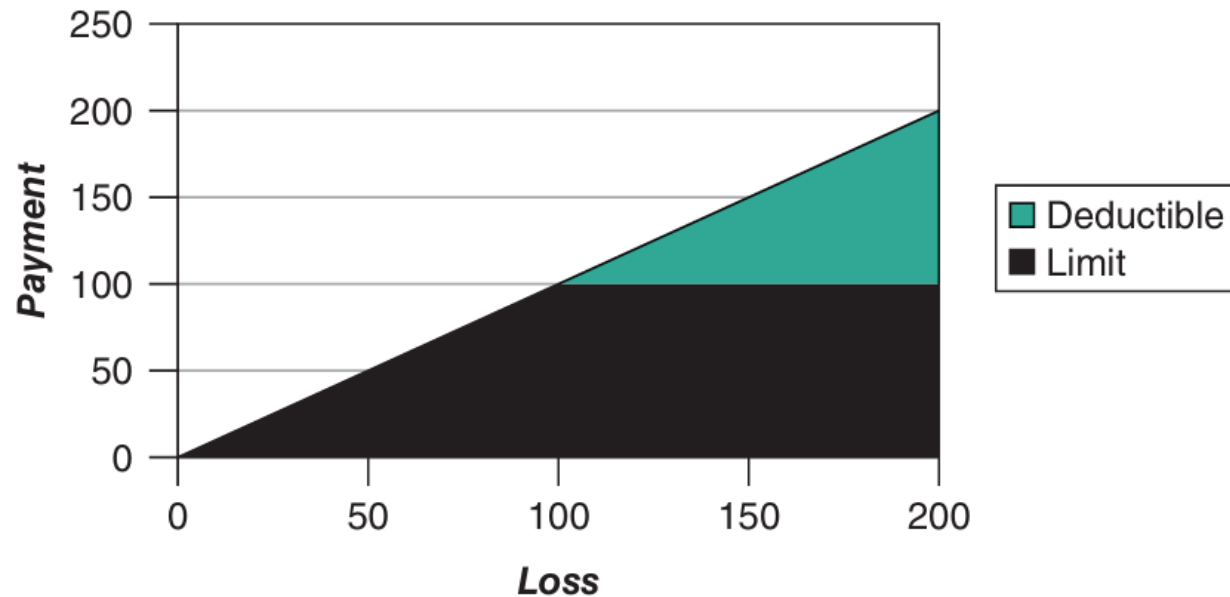
$$\text{Var}(Y^L) = S(d) \cdot E[(X - d)^2 \mid X > d] - [E(Y^L)]^2$$

# Limited Loss Variable

□ The **Limited Loss Variable**, often denoted as  $Y$ , represents the amount of loss that is "capped" at a certain limit. This variable could also be called the right censored variable.

That is

$$Y = X \wedge u = \min(X, u) = \begin{cases} X & X < u \\ u & X \geq u \end{cases}$$



□ Left Censored and Shifted Variable: calculate what the company pays after subtracting the "Deductible" (750).

□ Limited Loss Variable: calculate the portion of the loss the customer "retains" (the average cost of the deductible to the policyholder).

# Claim Modification

## Example

An automobile insurance policy with no coverage modifications has the following possible losses, with probabilities in parentheses:

|          |     |     |       |       |        |
|----------|-----|-----|-------|-------|--------|
| $X$      | 100 | 500 | 1,000 | 2,500 | 10,000 |
| $P_X(x)$ | 0.4 | 0.2 | 0.2   | 0.1   | 0.1    |

Determine the probability mass functions and expected values for the **excess loss**, left censored and shifted and limited loss variables, where the deductible is set at 750.

## Solution

The **excess loss** variable is a modified random variable  $Y^P = X - 750 | X > 750$

$$P(Y^P = y) = \frac{P(X - 750 \text{ and } X > 750)}{P(X > 750)}$$

|          |                         |                          |                          |
|----------|-------------------------|--------------------------|--------------------------|
| $Y^L$    | $1,000 - 750 = 250$     | $2,500 - 750 = 1,750$    | $10,000 - 750 = 9,250$   |
| $P_Y(y)$ | $\frac{0.2}{0.4} = 0.5$ | $\frac{0.1}{0.4} = 0.25$ | $\frac{0.1}{0.4} = 0.25$ |

$$e(d) = 250 \times 0.5 + 1,750 \times 0.25 + 9,250 \times 0.25 = 2,875.$$

# Claim Modification

---

## Example

An automobile insurance policy with no coverage modifications has the following possible losses, with probabilities in parentheses:

|          |     |     |       |       |        |
|----------|-----|-----|-------|-------|--------|
| $X$      | 100 | 500 | 1,000 | 2,500 | 10,000 |
| $P_X(x)$ | 0.4 | 0.2 | 0.2   | 0.1   | 0.1    |

Determine the probability mass functions and expected values for the excess loss, **left censored and shifted** and limited loss variables, where the deductible is set at 750.

## Solution

The **left censored and shifted** variable  $Y^P = \max(0, X - 750)$ , that is

$$P(X - 750 > 0) = P(X)$$

|          |     |     |     |       |       |
|----------|-----|-----|-----|-------|-------|
| $X$      | 0   | 0   | 250 | 1,750 | 9,250 |
| $P_X(x)$ | 0.4 | 0.2 | 0.2 | 0.1   | 0.1   |

$$E(P^P) = 0 \times 0.6 + 250 \times 0.2 + 1,750 \times 0.1 + 9,250 \times 0.1 = 1,150.$$

# Claim Modification

## Example

An automobile insurance policy with no coverage modifications has the following possible losses, with probabilities in parentheses:

|          |     |     |       |       |        |
|----------|-----|-----|-------|-------|--------|
| $X$      | 100 | 500 | 1,000 | 2,500 | 10,000 |
| $P_X(x)$ | 0.4 | 0.2 | 0.2   | 0.1   | 0.1    |

Determine the probability mass functions and expected values for the excess loss, left censored and shifted and **limited loss** variables, where the deductible is set at 750.

## Solution

The **limited loss variable**  $Y = \min(X, 750)$ , that is, values at or above 750 are assigned a value of 750

|          |     |     |     |     |     |
|----------|-----|-----|-----|-----|-----|
| $X$      | 100 | 500 | 750 | 750 | 750 |
| $P_X(x)$ | 0.4 | 0.2 | 0.2 | 0.1 | 0.1 |

$$E(Y) = 100 \times 0.4 + 500 \times 0.2 + 750 \times 0.4 = 440.$$

The expected value of the original random variable is

$$100(0.4) + 500(0.2) + 1,000(0.2) + 2,500(0.1) + 10,000(0.1) = 1,590$$

which is  $440 + 1,150$ .

# Claim Modification

---

## ❑ Excess Loss

- We "censor" (remove) all losses below 750. If you are looking at this data, you are only seeing the people who are "costing" the company money today.
- When a loss is large enough to trigger the insurance (i.e., it's over 750), the average payment is 2,875.
- The expected value is high because it only looks at the "expensive" end of the probability distribution.

## ❑ Left Censored and Shifted

- The amount of money the insurance company needs to set aside for **every person** they insure to cover the risk.
- If the average cost per payment is 1,150, the company must set premium more than that to cover their employees, rent, and profit.
- The expected value is lower than the average loss because many people (60% in your example) will get 0 from the insurer.

❑ The company expects to pay 2,875 on average, but they only have to do that 40% of the time. When you spread that cost across everyone (including the 60% who get nothing), the cost per person drops to 1,150.

❑ Therefore, what is important is not the variable being used but, rather, that it be used appropriately.

❑ The two values are linked by the likelihood of a payment being made

$$P(X > 750) = 0.4 \Rightarrow 1,150 = 2,875 \times 0.4$$

# Claim Modification

---

## Example

A claim severity distribution is **exponential with mean 1000**. An insurance company will pay the amount of each claim in **excess of a deductible of 100**. Calculate the variance of the amount paid by the insurance company for one claim, **including the possibility that the amount paid is 0**.

## Solution

This scenario describes payment per loss random variable  $Y^L$

$$Y^L = (X - d)_+ = \max(0, X - d),$$

The exponential distribution is memoryless.

$$P(X > s + t \mid X > s) = P(X > t)$$

**This is a massive shortcut for SOA exams.**

If  $X \sim \text{Exp}(\lambda)$ , then the "excess loss"  $(X - d \mid X > d) \sim \text{Exp}(\lambda)$

$$E[(Y^L)^k] = \int_d^\infty (x - d)^k \lambda e^{-\lambda x} dx$$

substitute  $u = x - d$  (so  $x = u + d$  and  $dx = du$ ):

$$E[(Y^L)^k] = \int_0^\infty u^k \lambda e^{-\lambda(u+d)} du = \int_0^\infty u^k \lambda e^{-\lambda d} e^{-\lambda u} du = \underbrace{e^{-\lambda d}}_{s(d)} \cdot \underbrace{\int_0^\infty u^k \lambda e^{-\lambda u} du}_{E[X^k]}$$

# Claim Modification

---

That is

$$E[(Y^L)^k] = S(d) \cdot E[X^k]$$

but

$$E[X^k] = \int_0^{\infty} x^k \lambda e^{-\lambda x} dx$$

Let  $u = \lambda x$ , which implies  $x = u/\lambda$  and  $dx = du/\lambda$

$$E[X^k] = \int_0^{\infty} \left(\frac{u}{\lambda}\right)^k \lambda e^{-u} \frac{du}{\lambda} = \frac{1}{\lambda^k} \int_0^{\infty} u^k e^{-u} du$$

This integration is **Gamma Function**,  $\Gamma(k + 1)$ . For any integer  $k$ ,  $\Gamma(k + 1) = k!$ .

$$\mathbf{E[X^k]} = \frac{\mathbf{k!}}{\lambda^k} \Rightarrow E[(Y^L)^k] = e^{-\lambda d} \cdot \frac{k!}{\lambda^k}$$

But  $\lambda = \frac{1}{1000}$ , and  $d = 100$ , then

$$E[Y^L] = 1,000e^{-0.1} = 904.48,$$

$$E[(Y^L)^2] = \frac{2!}{\lambda^2} e^{-0.1} = 2,000,000e^{-0.1} = 1,809,675$$

$$\text{Var}(Y^L) = E[(Y^L)^2] - (E[Y^L])^2 = 990,940$$

# Claim Modification

---

## Example

Under a group insurance policy, an insurer agrees to pay 100% of the medical bills incurred during the year by employees of a small company, **up to a maximum total of 1 million Riyals**. The total amount of bills incurred,  $X$ , has pdf

$$f_X(x) = \frac{x(4-x)}{9}, 0 < x < 3.$$

where  $x$  is measured in millions. Calculate the total amount, in millions of Riyals, the insurer would expect to pay under this policy.

## Solution

let  $X$ : total amount of claims paid. Insurer pays everything up to a limit  $u$ , then we use the **limited loss variable**

$$Y = X \wedge u = \begin{cases} X & X < u \\ u & X \geq u \end{cases}$$

$$E[Y] = \int_0^u x f_X(x) dx + \int_u^\infty u f_X(x) dx,$$

For  $u = 1$

$$E[Y] = \int_0^1 \frac{x^2(4-x)}{9} dx + \int_1^3 \frac{x^2(4-x)}{9} dx = 0.935 \text{ million}$$

# Claim Modification

---

## Example

Losses have a density function

$$f_X(x) = \begin{cases} 0.02x & 0 < x < 10 \\ 0 & \text{otherwise} \end{cases}$$

The insurance has an ordinary deductible of 4 per loss.  $Y^P$  is the claim payment per payment random variable. Calculate  $E(Y^P)$

## Solution

We define  $Y^P$  as follows

$$Y^P = \begin{cases} \text{undefined} & X < 4 \\ X - 4 & X > 4 \end{cases}$$

$$E[Y^P] = E[X - d \mid X > d] = \frac{E[X - d]}{P[X > d]} = \frac{\int_4^{10} (x - 4) 0.02x \, dx}{\int_4^{10} 0.02x \, dx}$$

## Class Work

# Generating functions and sums of random variables

---

## Example

Consider a portfolio of insurance risks covered by insurance policies each has a gamma distribution with the same value of  $\theta$ . If each policy is independent of the other, show that the total claims paid by the insurance company has a gamma distribution.

## Solution

Let  $X_i$  represent the claim amount for the  $i^{th}$  policy, then  $X_i \sim \text{Gamma}(\alpha_i, \theta)$

The MGF of  $X_i$  is defined as:

$$M_X(t) = (1 - \theta t)^{-\alpha_i} \quad \text{for } t < \frac{1}{\theta}$$

Let  $S$  be the total claims for a portfolio of  $n$  independent policies:

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

Since the policies are independent, the MGF of the sum is the product of the individual MGFs

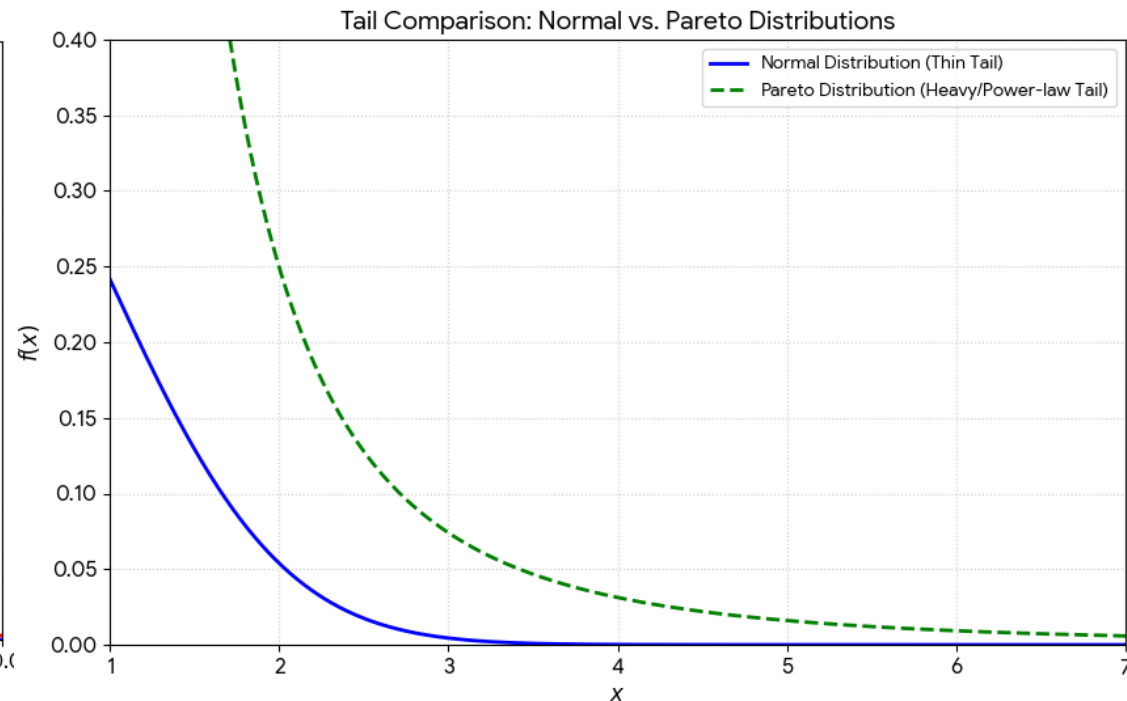
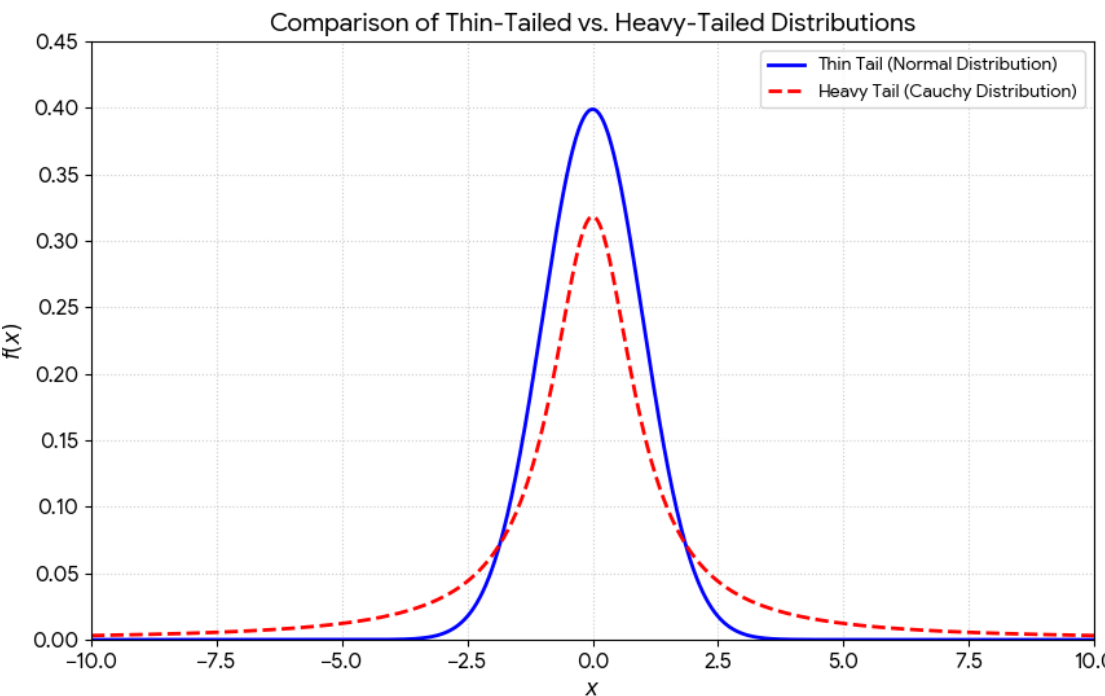
$$\begin{aligned} M_S(t) &= E[e^{tS}] = E[e^{t(X_1 + X_2 + \dots + X_n)}] = M_{X_1}(t) \cdot \dots \cdot M_{X_n}(t) \\ &= \prod_{i=1}^n (1 - \theta t)^{-\alpha_i} = (1 - \theta t)^{-\sum \alpha_i}. \end{aligned}$$

Therefore, the total claims  $S$  follow a Gamma distribution with parameters  $\alpha = \sum_{i=1}^n \alpha_i$  and  $\theta$

# **CHARACTERISTICS OF ACTUARIAL MODELS**

# Tails of distributions

- ❑ Tail of a distribution is the portion of the distribution corresponding to **large or small values** of the random variable.
- ❑ Actuary is interested in the **right tail**, that is the distribution of **large values** since they have the greatest **impact on the total of losses**.
  - **Thin Tail** (Normal): future is predictable and extreme surprises are almost impossible.
  - **Heavy Tail** (Pareto): further you move in time, a higher chance of an event that costs millions.
- ❑ Heavy tails indicate that **catastrophic losses are not only possible but expected**.
- ❑ Heavy tails necessitate **Reinsurance**.
- ❑ Existence of **all positive moments** indicates a thin tail,
- ❑ Existence of **non-positive moments** indicates a heavy right tail.



# Tails of distributions

---

## □ The Three Stages of the Tail

- **Reporting Lag** (creates the Unknown): It might take years for an injury to be discovered. For example, a person might be exposed to a chemical today but not get sick for 10 years.
- **Development Lag** (creates the Severity): Once the claim is reported, the cost can grow. A claim that looks like it will cost 10,000 might turn into 1,000,000 if the person's health worsens or if a court awards high damages.
- **Legal Lag** (creates the Infinite Tail): High-value lawsuits often go through years of discovery, depositions, and trials before a single rial or dollar is paid.

## Example

If  $X \sim \text{gamma}(\alpha, \theta)$  with  $\alpha, \theta > 0$ , show that this distribution has thin tail

## Solution

$$\mu'_k = E[X^k] = \int_0^\infty x^k f(x; \alpha, \theta) dx = \int_0^\infty x^k \left( \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta} \right) dx,$$

Let  $y = \frac{x}{\theta} \rightarrow x = \theta y \rightarrow dx = \theta dy$

$$\mu'_k = \int_0^\infty (\theta y)^k \left( \frac{1}{\Gamma(\alpha)\theta^\alpha} (\theta y)^{\alpha-1} e^{-y} \right) \theta dy = \frac{\theta^k}{\Gamma(\alpha)} \int_0^\infty y^{k+\alpha-1} e^{-y} dy = \frac{\Gamma(\alpha+k)\theta^k}{\Gamma(\alpha)}.$$

□ Since  $\theta^k$ ,  $\Gamma(\alpha + k)$  and  $\Gamma(\alpha)$  finite, positive constants, the result must be finite

$$\mu'_k < \infty \quad \text{for all } k > 0$$

**Exercise:** determine the nature of the tail for Weibull and Pareto distributions.

# The Role of Parameters

---

## □ Parametric Distribution:

- Set of distribution functions each member of which is determined by specifying **one or more values called parameters**. The number of parameters is fixed and finite.
- parameters specify how much information is needed to describe the model.
- For example: normal distribution with parameters  $\mu$  and  $\sigma^2$ .

## □ The number of parameters needed to describe the model indicate the **complexity of a distribution**.

## □ The benefit of models with small number of parameters (simple models)

- Each item can be determined more accurately.
- Stable across time and across settings. Small changes in parameters reflects inflation or similar situations.

## □ In the other hand complex models also have advantages

- More closely match reality.
- Match irregularities in the data.

## □ For actuarial modeling, it is convenient if the distribution is unchanged when the random variable is multiplied by a constant.

- The most common uses for this phenomenon are to model the effect of inflation and to accommodate a change in the monetary unit.

## □ Scale Distribution

- when a random variable from that set of distributions is **multiplied by a positive constant**, the resulting random variable is also in that set of distributions.

# The Role of Parameters

## Example

Demonstrate that the exponential distribution is a scale distribution.

## Solution

Let  $X$  be a random variable following exponential distribution with rate parameter  $\theta$ ,  
The distribution function for  $X$  is

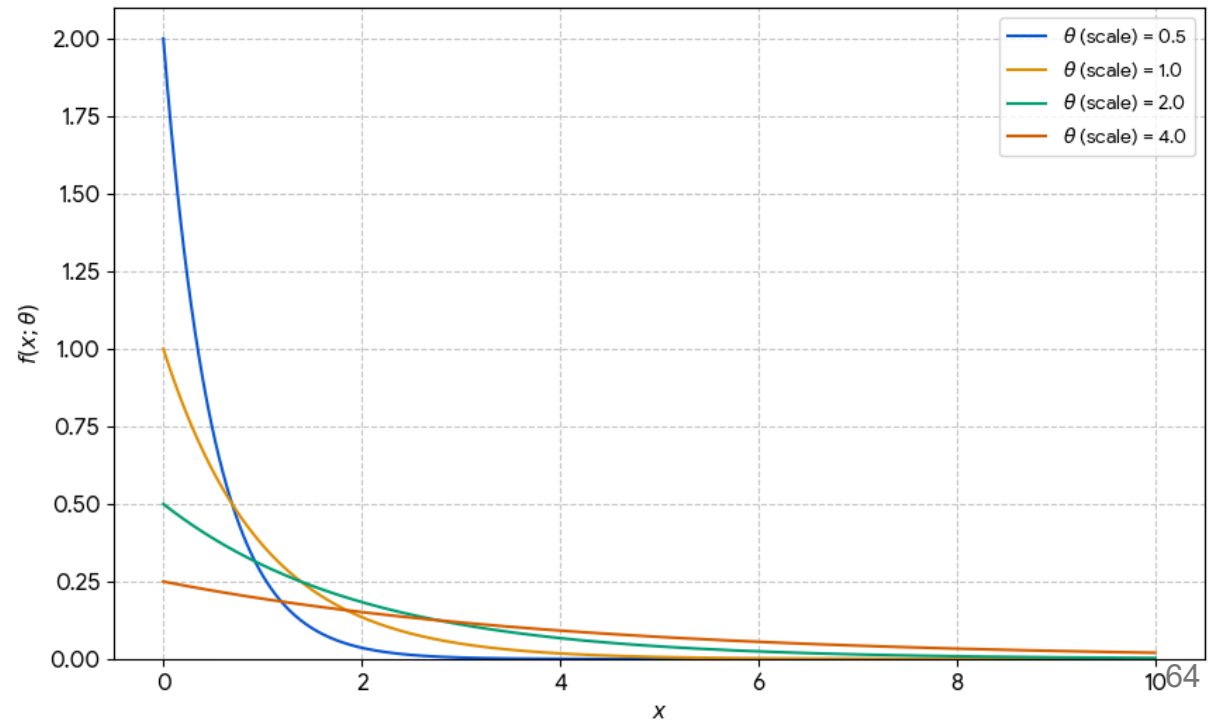
$$F_X(x) = 1 - e^{-\frac{x}{\theta}},$$

define a random variable  $Y = cX$ , where  $c > 0$  is a constant scaling factor, then

$$F_Y(y) = P[Y \leq y] = P[cX \leq y] = P\left[X \leq \frac{y}{c}\right] = 1 - e^{-\frac{y}{c\theta}}.$$

The resulting distribution function for  $Y$  is still in the exact same exponential form, just with a new scale parameter  $c\theta$ .

The shape remains fundamentally the same, a decaying exponential, varying only by the horizontal stretch factor  $\theta$ .



# The Role of Parameters

---

## □ Scale Parameter

A parameter for a that meets two conditions.

when a member of the scale distribution is **multiplied by a positive constant**,

1. the **scale parameter is multiplied by the same constant**.
2. **all other parameters are unchanged**.

## Example

Demonstrate that the gamma distribution has a scale parameter

## Solution

The Gamma distribution is essentially a generalization of the exponential distribution, defined by two parameters: shape ( $\alpha$ ) and scale ( $\theta$ ).

Let  $X \sim \text{Gamma}(\alpha, \theta)$ , then the probability density function is

$$f_X(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha) \theta^\alpha} \gamma\left(\alpha, \frac{x}{\theta}\right),$$

Where  $\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt$ : Lower Incomplete Gamma Function.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(cX \leq y) = P\left(X \leq \frac{y}{c}\right) = F_X\left(\frac{y}{c}\right) \\ &= \frac{1}{\Gamma(\alpha) \theta^\alpha} \gamma\left(\alpha, \frac{y}{c\theta}\right) \end{aligned}$$

$\theta$  is a scale parameter since gamma distribution has parameters  $\alpha$  and  $c\theta$ .

# Parametric Distribution Families

---

- ❑ A slightly **more complex** version of a parametric distribution is one in which the **number of parameters is finite** but **not fixed** in advance.
- ❑ Parametric Distribution Family
  - A set of parametric distributions that are **related** in some meaningful way
  - members of the family are obtained by **setting** one or more **parameters** from the specified distribution **equal to a preset value** or to **each other**.

## Example

Explain how beta distribution forms a parametric distribution family

## Solution

For the Beta distribution, we define it on the interval  $[0, 1]$  using the PDF:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)},$$

where  $B(\alpha, \beta)$  is the Beta function  $B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ .

By adjusting  $\alpha$  and  $\beta$ , we can model completely different behaviors:

**Uniform** :  $\alpha = 1$  and  $\beta = 1$ , Every outcome between 0 and 1 is equally likely.

**Power-Law**: one parameter is 1 and the other is  $> 1$ , the distribution becomes a (triangular shape), leaning toward 0 or 1.

**Bell-Shaped**: both  $\alpha, \beta > 1$ , the distribution has a peak (mode) in the middle. If  $\alpha = \beta$ , it is perfectly symmetric.

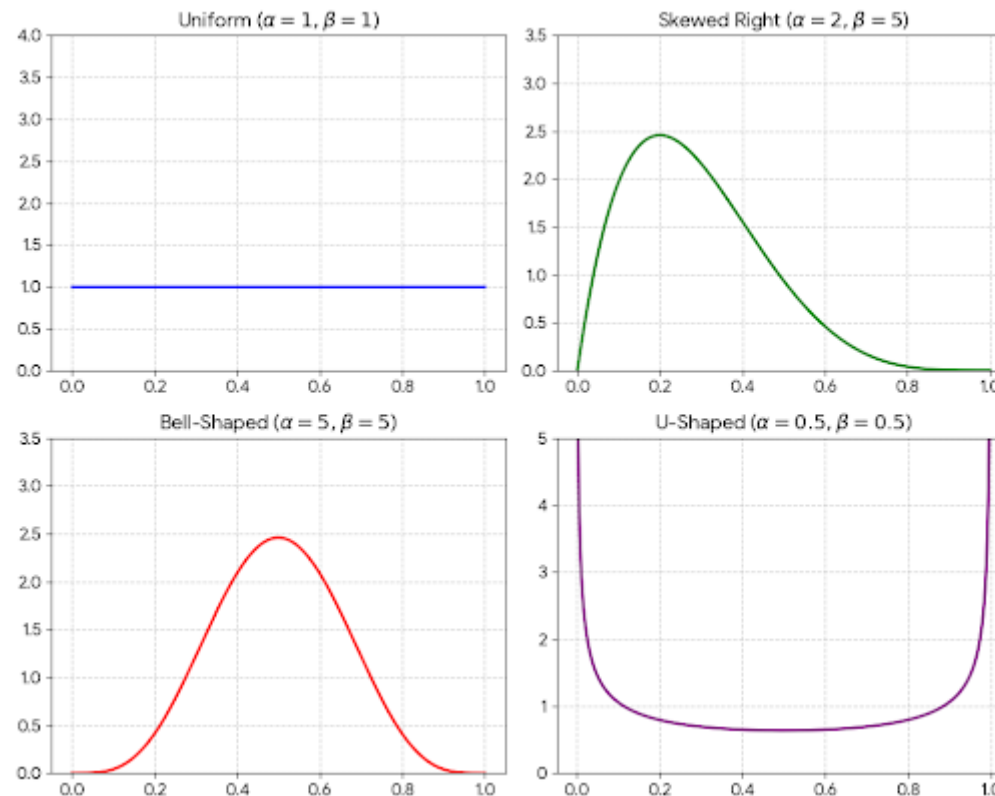
**U-Shaped**: both  $\alpha, \beta < 1$ , the density "explodes" at the edges (0 and 1),

# Beta Distribution parametric Family

$\alpha$ : Often thought of as "counts of successes + 1." Increasing  $\alpha$  pulls the distribution toward 1.

$\beta$ : Often thought of as "counts of failures + 1." Increasing  $\beta$  pulls the distribution toward 0.

□ By varying these two values, you can navigate through every possible "member" of this distribution family to fit your specific data.



# Finite Mixture Distributions

---

- ❑ In actuarial science, a mixture distribution is the probability distribution of a random variable that is derived from a **collection of other random variables**.
  - A randomly selected dental claim may be from a check-up, filling, repair, or surgical procedure.
  - A motor insurance portfolio may have 70% of the group "Safe Drivers" and 30% "High-Risk Drivers" (30%). A single Poisson distribution won't fit the data; instead, so using a mixture of two Poisson distributions with weights 0.7 and 0.3 will be more accurate.
- ❑ Gamma distribution fits the small claims, but it fails to capture the rare, massive claims. Actuaries might use a mixture of a **Gamma** (for the body) and a **Pareto** (for the tail).

## ❑ ***k*-Point Mixture**

A random variable  $Y$  is a  $k$ -point mixture of the random variables  $X_1, X_2, \dots, X_k$ , with distribution function  $F(x)$ , as the weighted sum of the individual distributions:

$$F(x) = \sum_{i=1}^k w_i F_i(x)$$

$w_i$  (Mixing Weights): The probability or proportion of the population belonging to subgroup  $i$ , such that  $w_i > 0$  and  $\sum_{i=1}^k w_i = 1$ .

$F_i(x)$  (Component Distributions): The individual distributions being mixed (e.g., Pareto, Exponential, or Gamma).

# Finite Mixture Distributions

---

## Example

For models involving **general liability** (GL) insurance, actuaries at the **Insurance Services Office** (ISO) once considered a **mixture of two Pareto distributions**.

$$F(x) = w \left[ 1 - \left( \frac{\theta_1}{x + \theta_1} \right)^{\alpha_1} \right] + (1 - w) \left[ 1 - \left( \frac{\theta_2}{x + \theta_2} \right)^{\alpha_2} \right],$$

This model contains a total of **five parameters**.

$w$ : The mixing weight

$\alpha_1$  and  $\alpha_2$ : The shape of the distributions (tail heaviness)

$\theta_1$  and  $\theta_2$ : The scale of the distributions.

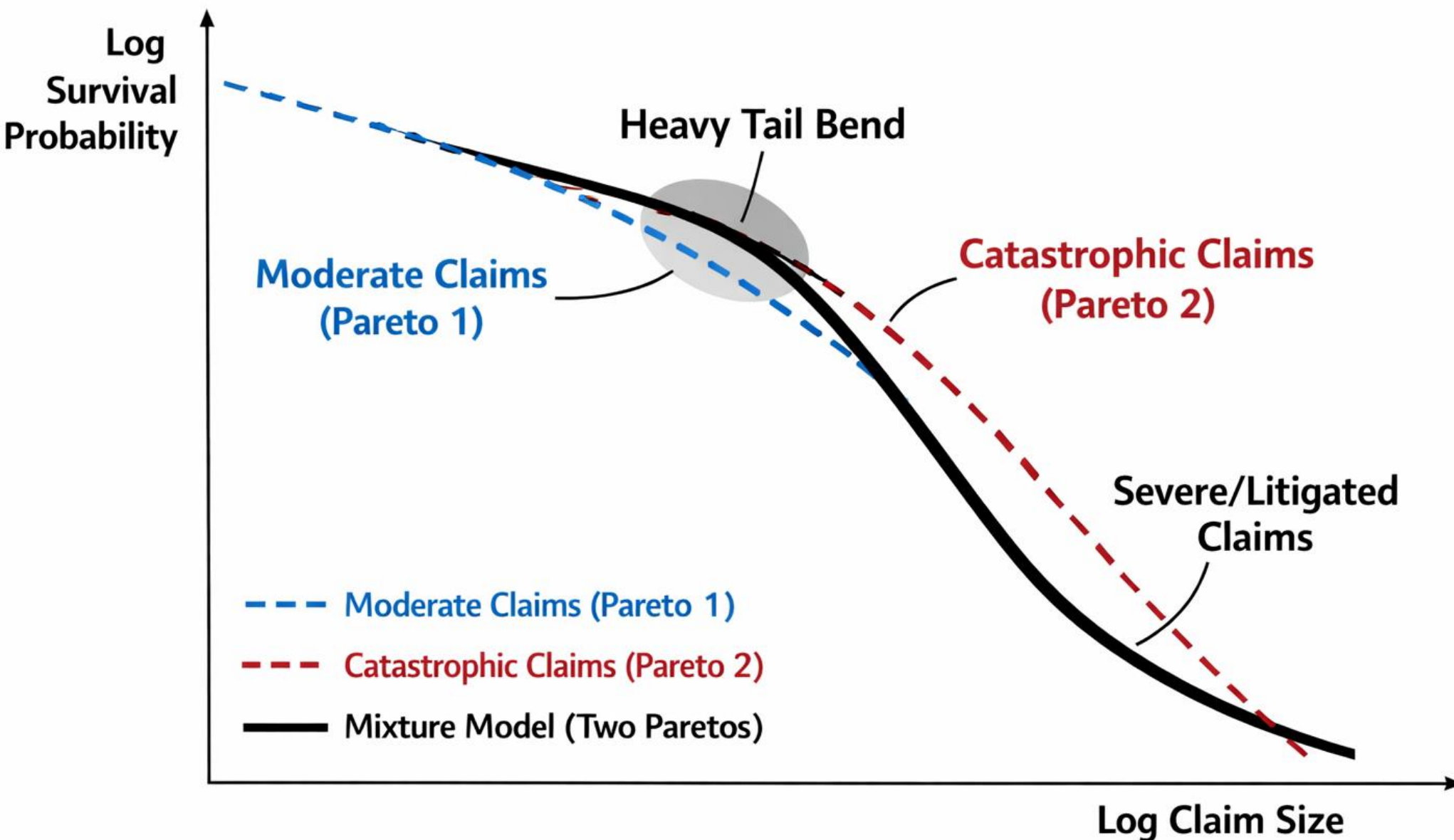
□ They decided that **five parameters were not necessary** and decided **reduce the "degrees of freedom"** the model by creating a **mathematical relationship between the two shape parameters** ( $\alpha_1$  and  $\alpha_2$ ), and selected the distribution with distribution function

$$F(x) = w \left[ 1 - \left( \frac{\theta_1}{x + \theta_1} \right)^{\alpha} \right] + (1 - w) \left[ 1 - \left( \frac{\theta_2}{x + \theta_2} \right)^{\alpha+2} \right],$$

□ The second distribution places more probability on smaller values. This might be a model for frequent, small claims

□ The first distribution covers large but infrequent claim.

# Mixture of Pareto Distributions for General Liability Claims



*Captures Both Routine Claims and Extreme Losses.*

# Finite Mixture Distributions

---

□ Suppose we do not know **how many distributions should be in the mixture**. Then the **value of  $k$  becomes a parameter**, as indicated in the following definition.

## □ Variable-Component Mixture Distribution

It has a distribution function that can be written as

$$F(x) = \sum_{i=1}^K w_i F_i(x), w_i > 0 \text{ and } \sum_{i=1}^K w_i = 1 \text{ and } K = 2, 3, \dots$$

□ When all of the components have the **same parametric** distribution (but **different parameters**), the resulting distribution is called a "**variable mixture of distribution  $gs$** ", where  $g$  stands for the name of the **component distribution**.

# Finite Mixture Distributions

---

## Example

Determine the distribution, **density**, and **hazard rate functions** for the variable mixture of exponentials distribution.

## Solution

The distribution function  $F(x)$  is the weighted sum of the individual exponential DFs.

For an exponential distribution,  $F_i(x) = 1 - e^{-\lambda_i x}$ .

$$F(x) = \sum_{i=1}^K w_i (1 - e^{-\lambda_i x}).$$

Since  $\sum w_i = 1$ , this can be simplified to:

$$F(x) = 1 - \sum_{i=1}^K w_i e^{-\lambda_i x}, \quad x \geq 0.$$

The density function  $f(x)$  is the derivative of the CDF,  $F'(x)$ . The mixture density is:

$$f(x) = \sum_{i=1}^K w_i \lambda_i e^{-\lambda_i x}, \quad x \geq 0.$$

The hazard rate  $h(x)$  represents the instantaneous rate of failure at time  $x$ , given survival up to that time. It is defined as:

$$h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1-F(x)} = \frac{\sum_{i=1}^K w_i \lambda_i e^{-\lambda_i x}}{\sum_{i=1}^K w_i e^{-\lambda_i x}}.$$

when

$K = 2$ : three parameters  $(a_1, \lambda_1, \lambda_2)$ , noting that  $a_2$  gets its value as  $a_1$  is set,

$K = 4$ : there are seven parameters.

# Finite Mixture Distributions

---

## Example

Illustrate how a two-point mixture of gamma variables can create a bimodal distribution. Consider a 50-50 mixture of two gamma distributions. One has parameters  $\alpha = 4$  and  $\theta = 7$ , for a mode of  $21=7(4-1)$  and the other has parameters  $\alpha = 15$  and  $\theta = 7$ , for a mode of  $98=7(15-1)$ .

## Solution

A Gamma distribution is defined by a **shape parameter**  $\alpha$  and a **scale parameter**  $\theta$ . we have two distinct groups, both using the same scale ( $\theta = 7$ ) but different shapes:

Component 1 (The "Lower" Peak)

Parameters:  $\alpha_1 = 4, \theta = 7$

Mode:  $(4 - 1) \times 7 = 21$

Component 2 (The "Higher" Peak):

Parameters:  $\alpha_2 = 15, \theta = 7$

Mode:  $(15 - 1) \times 7 = 98$

The combined Probability Density Function (PDF) is:

$$f_{mix}(x) = 0.5 \cdot f(x; \alpha = 4, \theta = 7) + 0.5 \cdot f(x; \alpha = 15, \theta = 7)$$

Bimodality isn't guaranteed in every mixture; it depends on how much "overlap" exists between the two components.

The distance between the modes ( $98 - 21 = 77$ ) is large relative to the spread (variance) of the individual distributions.

# Finite Mixture Distributions

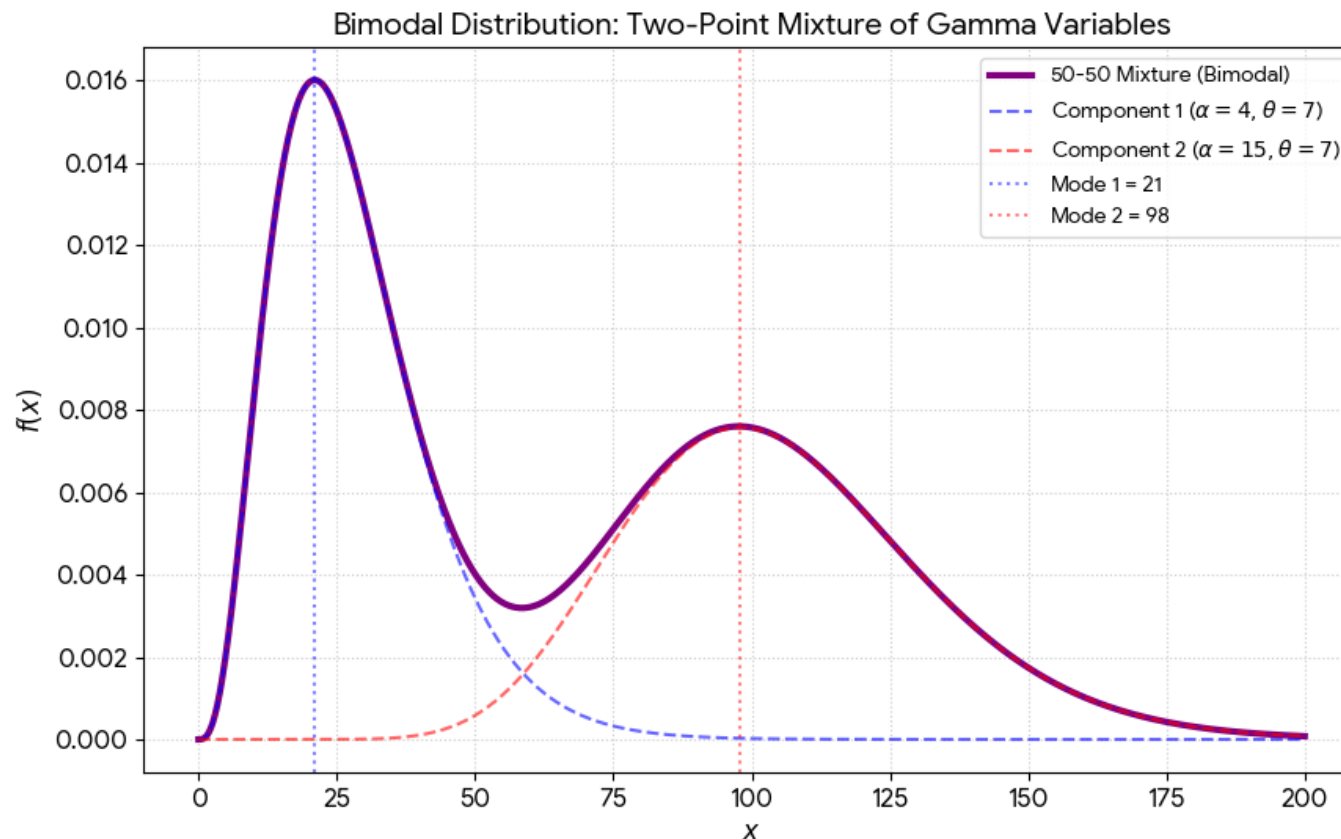
The mean of a mixture is simply the weighted average of the individual means.

$$E[X_1] = \alpha_1 \theta = 4 \times 7 = 28$$

$$E[X_2] = \alpha_2 \theta = 15 \times 7 = 105$$

$$E[X_{mix}] = (0.5 \times 28) + (0.5 \times 105) = \mathbf{66.5}$$

By using a mixture, the actuary avoids the error of using a single distribution that would put the "average" somewhere around 60—a value where very few actual claims occur.



# Data-dependent distributions

---

- ❑ In general, we assume the data follows a **specific, predefined shape (distribution)** with a specific parameters like a Normal or Poisson distributions.
- ❑ Data-dependent (AKA Non-parametric) **emerges from the data points**.
- ❑ The Empirical Distribution Function (EDF) is the interpretation of data.
  - Assigns a **probability of  $1/n$  to each of the  $n$  data points**.
  - Non-parametric baseline as it doesn't assume any specific shape.
  - Discrete and "jagged." Even if the data is smooth, it have gaps according to missing data.
- ❑ Kernel Smoothing (or Kernel Density Estimation, KDE) evolves the empirical distribution by replacing discrete data points with **smooth kernels**.
  - places a probability "cloud" over each point, then sum them together to get a smooth curve.
  - It acts as a locally **data-dependent estimator**. The shape of the final curve depends entirely on where the data points are clustered.
- ❑ the formula for KDE

$$\hat{f}_h(x) = \frac{1}{nb} \sum_{i=1}^n K\left(\frac{x - x_i}{b}\right)$$

$K$ : the kernel (the shape of the smoothening distribution).

$b$ : the bandwidth (the width of the kernel)

# Data-dependent distributions

---

## Example

Consider a sample of size 8 in which the observed data points were 3, 5, 6, 6, 6, 7, 7, and 10.

1. Find probability function for the empirical model,
2. Construct kernel smoothing model using the uniform kernel with bandwidth of 2.

## Solution

1. The empirical probability function

$$p(x) = \begin{cases} 0.125, & x = 3 \\ 0.125, & x = 5 \\ 0.375, & x = 6 \\ 0.25, & x = 7 \\ 0.125, & x = 10 \end{cases}$$

2. Uniform Kernel Definition:

A uniform kernel with bandwidth  $b$  spreads the probability  $1/n$  over the range  $[x_i - b, x_i + b]$ .

The height of each box is  $\frac{1}{n \times 2b} = \frac{1}{32}$  and box width  $2 \times b = 4$ .

# Data-dependent distributions

---

The final model is the sum of these rectangles. For any point  $x$ , the density is:

$$\hat{f}(x) = \frac{1}{32} \times (\text{number of observations within } [x - 2, x + 2])$$

$$\hat{f}(x) = \frac{1}{32} \sum_{i=1}^8 I(x_i - 2 \leq x \leq x_i + 2)$$

Where  $I(\cdot)$  is the indicator function.

The boundaries for every single data point

$x = 3$ : Covers  $[1, 5]$

$x = 5$ : Covers  $[3, 7]$

$x = 6$  ( $\times 3$ ): Covers  $[4, 8]$

$x = 7$  ( $\times 2$ ): Covers  $[5, 9]$

$x = 10$ : Covers  $[8, 12]$

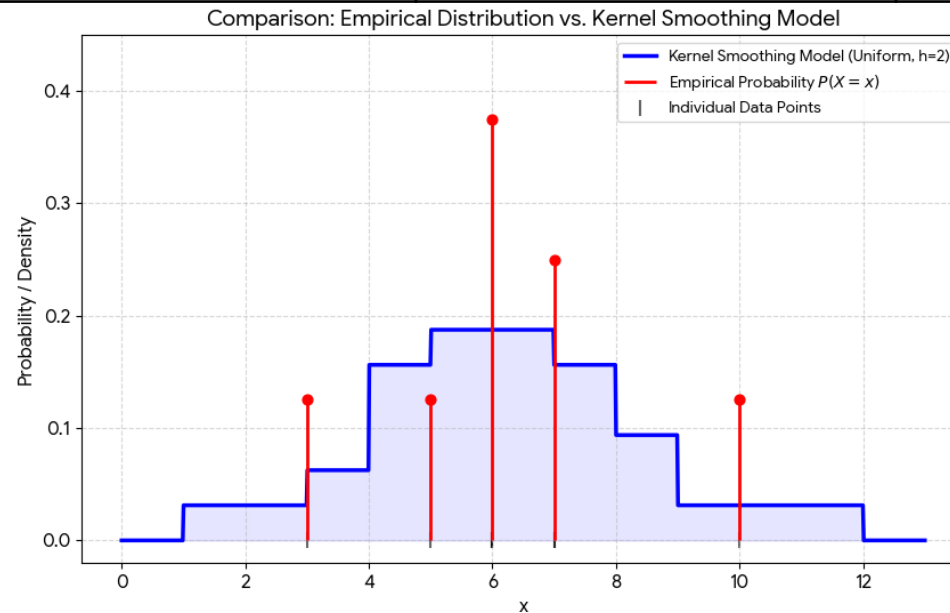
list all unique boundary values (without interval overlap) in order:

1, 3, 4, 5, 7, 8, 9, 12

These values define the intervals of our piecewise density function.

# Data-dependent distributions

| Interval | Data Points Covering this Range | Count | Density                |
|----------|---------------------------------|-------|------------------------|
| [1, 3)   | {3}                             | 1     | $1/32 \approx 0.03125$ |
| [3, 4)   | {3, 5}                          | 2     | $2/32 = 0.0625$        |
| [4, 5)   | {3, 5, 6, 6, 6}                 | 5     | $5/32 \approx 0.15625$ |
| [5, 7)   | {5, 6, 6, 6, 7, 7}              | 6     | $6/32 = 0.1875$        |
| [7, 8)   | {6, 6, 6, 7, 7}                 | 5     | $5/32 \approx 0.15625$ |
| [8, 9)   | {7, 7, 10}                      | 3     | $3/32 \approx 0.09375$ |
| [9, 12]  | {10}                            | 1     | $1/32 \approx 0.03125$ |



## Exercise:

Consider a sample with observed data points were 3, 5, 6, 6, 6, 7, 7, and 10.

1. Find probability function for the empirical model,
2. Construct kernel smoothing model using the exponential kernel  $K(u) = \frac{1}{2e^{-|u|}}$  with bandwidth of 2

# Creating New Distributions

---

□ A new distributions can be generated form any distribution using a **transformation** or any other **mathematical technique**.

## 1. Multiplication by a constant

- $X \sim f_X(x)$  and  $Y = c \times X$ , then  $f_Y(y) = \frac{1}{c} f_X\left(\frac{y}{c}\right)$ ,
- $c$  is a scale parameter/or the random variable.
- Equivalent to **applying inflation uniformly** across all losses and is known as a **change of scale**

## 2. Raising to a power

- $X \sim f_X(x)$  and  $Y = X^c$ , then  $f_Y(y) = \left| \frac{1}{c} y^{\frac{1}{c}-1} \right| f_X(y^{1/c})$ ,
- $c > 0$ ,  $Y$ : transformed;  $c = -1$ ,  $Y$ : inverse;  $c < 0$  ( $c \neq -1$ ),  $Y$ : inverse transformed.
- Map skewed claim (**outlier**) amounts into a symmetric, often approximately normal, distribution ( $0 < c < 1$ ), and **reduce the effect of volatility** so predictions are equally accurate for small and large claims, also helps to calculate the probability of "**once in a century**" events.

## 3. Exponentiation

- $X \sim f_X(x)$  and  $Y = e^X$ , then  $f_Y(y) = \frac{1}{y} f_X(\ln(y))$
- exponentiation is often used to model the underlying process of growth or compounding.

# Mixed Distributions (Infinite Mixture Distribution)

---

- ❑ In case of using a single distribution (no mixture) to model the risk, we might underprice the "tail risk" (the 1-in-100-year storm).

$$f(x; \lambda) = P(X = x | \lambda)$$

- ❑ An infinite mixture (a few distinct groups) naturally creates a heavier tail, meaning the model "expects the unexpected" more accurately than a simple model would.

$$f(x) = \sum_{i=1}^k w_i \cdot f(x; \lambda_i); \sum w_i = 1.$$

- ❑ Mixing a finite number of distribution, we created a weighted mean of all distributions, for example, A portfolio of safe driving group ( $w_1 = 80\%$ ) and high-risk drivers' group ( $w_2 = 20\%$ ).
- ❑ In this model we consider all drivers in each group has the same distribution with unique parameter. But in the real world, **everyone is different**.
- ❑ Instead of trying to find **one fixed  $\lambda$  (claim frequency)** that fits everyone, we treat the **parameter as a random variable**, denoted by the uppercase  $\Lambda$ .

That is

Every individual in your portfolio is assigned a **realization** of  $\Lambda$ .  
As an actuary, you cannot see their specific  $\lambda$ ,  
but you know the distribution of  $\Lambda$  across the whole population.

# Mixed Distribution

---

- ❑ Let  $X$  be a continuous random variable with pdf  $f_X(x|\lambda)$ , which depends on the parameter  $\lambda$ .
- ❑ We allow  $\lambda$  to be the realization of a random variable  $\Lambda$  with support  $\Omega_\Lambda$  and pdf  $f_\Lambda(\lambda|\theta)$ , where  $\theta$  is the parameter determining the distribution of  $\Lambda$ , sometimes called the hyperparameter.
- ❑ We define two layers of probability:
  1. Conditional Distribution:  $X|\Lambda = \lambda \sim f_{(X|\Lambda)}(x|\lambda)$ .
    - Probability of individual risk depends on  $\lambda$  which is chosen from population distribution  $f_\Lambda(\lambda)$ .
    - Therefore, the probability of  $X$  is conditioned on the value of  $\Lambda$  as a **prior state**.
    - $f_{(X|\Lambda)}(x|\lambda)$ : risk of an individual ( $x$ ) given their risk level ( $\lambda$ ).
  2. Mixing Distribution:  $\Lambda \sim f_\Lambda(\lambda; \theta)$ .
    - The distribution of those risk levels across the population, governed by hyperparameters  $\theta$ .
    - $\Lambda$  is the "Potential" (The Random Variable),  $\lambda$  is the "Identity" (The Realization).
- ❑ The **unconditional (marginal) probability** of an observation  $x$  for a randomly selected individual, we must account for every possible realization of  $\lambda$ . Since  $\Lambda$  is continuous, we use an integral for the mixture:

$$f_X(x) = \int_{\text{all } \lambda} f_{X|\Lambda}(x|\lambda) \cdot f_\Lambda(\lambda) d\lambda$$

# Mixed Distribution

---

□ **Distribution function** for  $\Lambda$ ,  $F_\Lambda(\lambda)$ , can be determined from

$$\begin{aligned} F_\Lambda(\lambda) &= P(\Lambda \leq x) = \int_0^x f_\Lambda(t) dt = \int_0^x \int f_{X|\Lambda}(x|\lambda) \cdot f_\Lambda(\lambda) d\lambda dt \\ &= \int \int_0^x f_{X|\Lambda}(x|\lambda) \cdot f_\Lambda(\lambda) d\lambda = \int F_{X|\Lambda}(x|\lambda) \cdot f_\Lambda(\lambda) d\lambda \end{aligned}$$

□ **Moments** of the mixture distribution can be found as follows

$$E[X^k] = \int \left( \int x^k f_{X|\Lambda}(x|\lambda) dx \right) f_\Lambda(\lambda) d\lambda = \int E[X^k|\lambda] f_\Lambda(\lambda) d\lambda = E_\Lambda [E[X^k|\lambda]]$$

□ **Mean** of the infinite mixture is the **expected value of the individual means**.

Mathematically, it is **the average of the realizations**.

$$E[X] = E_\Lambda[E[X|\Lambda]],$$

- The average number of claims is the average of the total risk parameters

□ **Variance** of a mixture is **not the average of the variances**

$$Var(X) = \underbrace{E_\Lambda[Var(X|\Lambda)]}_{\text{Process Variance}} + \underbrace{Var_\Lambda(E[X|\Lambda])}_{\text{Structural Variance}},$$

- Structural variance caused by the fact that  $\Lambda$  is a random variable and not a constant.

# Mixed Distribution

---

## Example

Let  $X|\Lambda$  have an exponential distribution with parameter  $1/\Lambda$ . Let  $\Lambda$  have a gamma distribution. Determine the **unconditional distribution** of  $X$

## Solution

Given a realization  $\Lambda = \lambda$ , the random variable  $X$  follows an Exponential distribution with parameter  $1/\lambda$ .  $X|\Lambda \sim \text{Exp}(1/\lambda)$ , the PDF is:

$$f_{X|\Lambda}(x|\lambda) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}x}, \quad x > 0$$

The parameter  $\Lambda \sim \text{Gamma}(\alpha, \beta)$

$$f_{\Lambda}(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{-(\alpha+1)} e^{-\beta/\lambda}, \quad \lambda > 0$$

The unconditional distribution of  $X$

$$\begin{aligned} f_X(x) &= \int_0^{\infty} f_{X|\Lambda}(x|\lambda) \cdot f_{\Lambda}(\lambda) d\lambda = \int_0^{\infty} \left( \frac{1}{\lambda} e^{-x/\lambda} \right) \left( \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{-(\alpha+1)} e^{-\beta/\lambda} \right) d\lambda \\ &= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} \lambda^{-(\alpha+2)} e^{-(x+\beta)/\lambda} d\lambda \end{aligned}$$

# Mixed Distribution

## Example

Let  $X|\Lambda$  have an exponential distribution with parameter  $1/\Lambda$ . Let  $\Lambda$  have a gamma distribution. Determine the **unconditional distribution**, of  $X$

## Solution

Substitute  $u = \frac{x+\beta}{\lambda} \Rightarrow \lambda = \frac{x+\beta}{u} \Rightarrow d\lambda = -\frac{x+\beta}{u^2} du$

if  $\lambda \rightarrow 0$ ,  $u = \frac{x+\beta}{\lambda} \rightarrow \infty$  and, if  $\lambda \rightarrow \infty$ ,  $u = \frac{x+\beta}{\lambda} \rightarrow 0$

$$\begin{aligned} \int_0^\infty \lambda^{-(\alpha+2)} e^{-(x+\beta)/\lambda} d\lambda &= \int_\infty^0 \left( \frac{x+\beta}{u} \right)^{-(\alpha+2)} e^{-u} \left( -\frac{x+\beta}{u^2} \right) du \\ &= (x+\beta)^{-(\alpha+1)} \int_0^\infty u^\alpha e^{-u} du = \frac{\Gamma(\alpha+1)}{(x+\beta)^{\alpha+1}}, \end{aligned}$$

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

That is:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{(x+\beta)^{\alpha+1}} = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}$$

Thus, the gamma-exponential mixture has a Pareto distribution.

## Exercise:

Show that gamma-Poisson mixture has a negative-binomial distribution.

# Frailty models

---

- ❑ The frailty is meant to quantify uncertainty associated with the hazard rate.
- ❑ In standard models, we assume a fixed hazard rate  $\lambda$ .
- ❑ In a Frailty Model, we treat the hazard itself as a Random Variable  $\Lambda > 0$ .
- ❑ Population is a continuous mixture of individuals with different risk levels.
- ❑ Define

- $X | \Lambda = \lambda$ : conditional hazard rate given  $\Lambda = \lambda$  as a conditional loss random variable.
- $\Lambda$ : The "Mixing Distribution" (The Frailty).

- ❑ Model the conditional hazard as a multiplicative effect on a baseline hazard  $a(x)$

$$h(x|\lambda) = \lambda \cdot a(x)$$

- This implies that if  $\lambda > 1$ , the individual is "more frail" (higher risk of loss) than the average.
- $a(x)$  is to be specified in a particular application.

- ❑ To find the **unconditional (observed) survival function**  $S_X(x)$ , we calculate the "weighted average" of all possible individual survival outcomes:

$$S_X(x) = \int_0^\infty S_{X|\Lambda}(x|\lambda) \cdot f_\Lambda(\lambda) d\lambda,$$

- ❑ If we assume a proportional hazard  $S_{X|\Lambda}(x|\lambda) = e^{-\lambda A(x)}$ , the integral becomes

$$S_X(x) = E[e^{-\Lambda A(x)}] = M_\Lambda[-A(x)],$$

Where,  $A(x) = \int_0^\infty a(t)dt$

# Frailty models

---

## Example

Let  $\Lambda$  have a gamma distribution and let  $X|\Lambda$  have a Weibull distribution with conditional survival function  $S_{x|\Lambda}(x|\lambda) = e^{-\lambda x^\gamma}$ . Determine the unconditional or marginal distribution of  $X$ .

## Solution

The Mixing Distribution (Frailty)  $\Lambda \sim \text{Gamma}(\alpha, \theta) \Rightarrow f_\Lambda(\lambda) = \frac{1}{\Gamma(\alpha)\theta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\theta}$ ,

The Conditional Distribution  $X|\Lambda = \lambda \sim \text{Weibull}(\lambda, \gamma)$

The conditional survival function is  $S_{X|\Lambda}(x|\lambda) = e^{-\lambda x^\gamma}$

To find the marginal survival function  $S_X(x)$ , we integrate over all values of  $\lambda$

$$S_X(x) = \int_0^\infty S_{X|\Lambda}(x|\lambda) f_\Lambda(\lambda) d\lambda = \int_0^\infty e^{-\lambda x^\gamma} f_\Lambda(\lambda) d\lambda = M_\Lambda(t) = (1 - \lambda x^\gamma)^{-\alpha}$$

This is a Burr distribution with the usual parameter  $\theta$  replaced by  $\theta^{-\frac{1}{\gamma}}$ .

Note that when  $\gamma = 1$ , this is an exponential mixture which is a Pareto distribution.

□ Mixing tends to create heavy-tailed distributions, and in particular a mixture of distributions that all have decreasing hazard rates also has a decreasing hazard rate.

# Splicing

---

□ Splicing is another powerful way to build flexible loss distributions, but it takes a **piecewise** approach rather than the **mixing** approach in that **one model** governs the behavior of losses in **some interval** of possible losses while **other models** cover the **other intervals**.

□ A  $k$ -component spliced model divides the loss domain  $(0, \infty)$  into  $k$  non-overlapping intervals using  $k - 1$  breakpoints (thresholds):  $0 = c_0 < c_1 < c_2 < \dots < c_k = \infty$ . The density function  $f_X(x)$  is defined piecewise:

$$f_X(x) = \begin{cases} a_1 f_1(x) & c_0 < x \leq c_1 \\ a_2 f_2(x) & c_1 < x \leq c_2 \\ \vdots & \\ a_k f_k(x) & c_{k-1} < x < c_k \end{cases}$$

each  $f_j(x)$  must be legitimate density function with all probability on  $(c_{j-1}, c_j)$ .

$$\sum_{j=1}^k \int_{c_{j-1}}^{c_j} a_j f_j(x) dx = 1$$

Also,  $a_1 + \dots + a_k = 1$

# Splicing

## Example

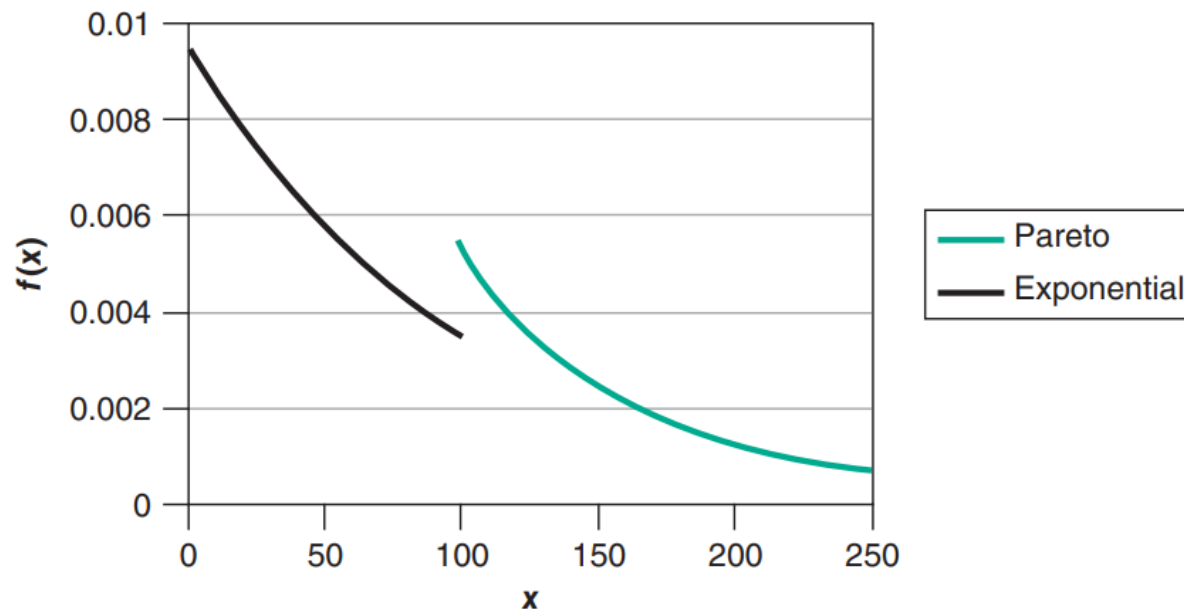
Create a two-component spliced model using an exponential distribution from 0 to  $c$  and a Pareto distribution (using  $\gamma$  in place of  $\theta$ ) from  $c$  to  $\infty$ .

## Solution

Component 1 (Exponential):  $f_1(x) = \lambda e^{-\lambda x}, 0 < x \leq c$

Component 2 (Pareto):  $f_2(x) = \frac{\alpha \gamma^\alpha}{x^{\alpha+1}}, x > c$

$$f_X(x) = \begin{cases} w \frac{f_1(x)}{F_1(c)} & 0 < x \leq c \\ (1-w) \frac{f_2(x)}{1-F_2(c)} & x > c \end{cases} = \begin{cases} w \cdot \frac{\lambda e^{-\lambda x}}{1-e^{-\lambda c}} & 0 < x \leq c \\ (1-w) \cdot \frac{\alpha(c+\gamma)^\alpha}{(x+\gamma)^{\alpha+1}} & x > c \end{cases},$$



# Example (Exercises 5.2.7-2)

One hundred observed claims in a given year arranged as follows:

| Interval | 0-300 | 300-350 | 350-400   | 400-450 | 450-500 | 500-600 | >600 |
|----------|-------|---------|-----------|---------|---------|---------|------|
| Claims   | 42    | 3       | 5         | 5       | 0       | 5       | 40   |
| Selected | Below |         | UNCERTAIN | Above   |         |         |      |

Based on this empirical distribution, determine a range for the probability that a claim exceeds 500 after three years if all claims are inflated by 10% per year.

## Solution

Let  $X$  represent a claim amount in the given year and  $Y$  represents the same claim amount after three years.

The relationship between  $X$  and  $Y$  is:

$$Y = X \times (1 + 0.10)^3 = X \times (1.1)^3 = 1.331X$$

The probability that a claim after three years exceeds 500:

$$P(Y > 500) = (1.331X > 500) = P\left(X > \frac{500}{1.331}\right) \approx P(X > 375.657).$$

Intervals that satisfies  $X > 375.6$  with the uncertainty in 350-400, we have:

$$\begin{aligned} P(X > 400) < P(X > 375.657) < P(X > 350) \\ \frac{5 + 0 + 5 + 40}{100} = 0.5 < P(X > 375.657) < 0.55 = \frac{5 + 5 + 0 + 5 + 40}{100} \end{aligned}$$

## Example (Exercises 5.2.7-19)

---

Losses in a given year follow the density function  $f(x) = 3x^{-4}, x > 1$ , where  $x$  is the loss in millions of dollars. Inflation of 10% impacts all claims uniformly in the next year. Determine the CDF of losses for the next year and use it to determine the probability that a 1994 loss exceeds 2,200,000.

### Solution

Obtain the CDF by integrating the density:

$$F_X(x) = \int_1^x 3t^{-4} dt = [-t^{-3}]_1^x = -x^{-3} - (-1^{-3}) = 1 - x^{-3}, \quad x > 1$$

Let  $X$  and  $Y$  represent the loss in that year and the year after it. Since inflation of 10% impacts all claims uniformly:

$$Y = (1 + 0.10)X = 1.1X$$

The CDF for next year losses,  $F_Y(y)$ , is derived using the transformation method:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(1.1X \leq y) = P\left(X \leq \frac{y}{1.1}\right) = F_X\left(\frac{y}{1.1}\right) \\ &= 1 - \left(\frac{y}{1.1}\right)^{-3} = 1 - \frac{1.1^3}{y^3} = 1 - \frac{1.331}{y^3}, \quad y > 1.1 \end{aligned}$$

That is

$$P(Y > 2.2) = 1 - P(Y < 2.2) = F_Y(2.2) = \frac{1.331}{2.2^3} = 0.125$$

## Exercise

---

The distribution of an unlimited claim-size random variable  $X$  has the exponential CDF

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-0.01x}, & x \geq 0 \end{cases}$$

Find the CDF for the following variables

- a) The variable  $Y$  takes on the value 0 with probability 0.25.
- b) The variable  $Y$  is limited above by the value 200.
- c) The variable  $Y$  has both modifications described in parts (a) and (b): it takes on the value 0 with probability  $P(X = 0) = 0.25$  and is limited above by the value 200.

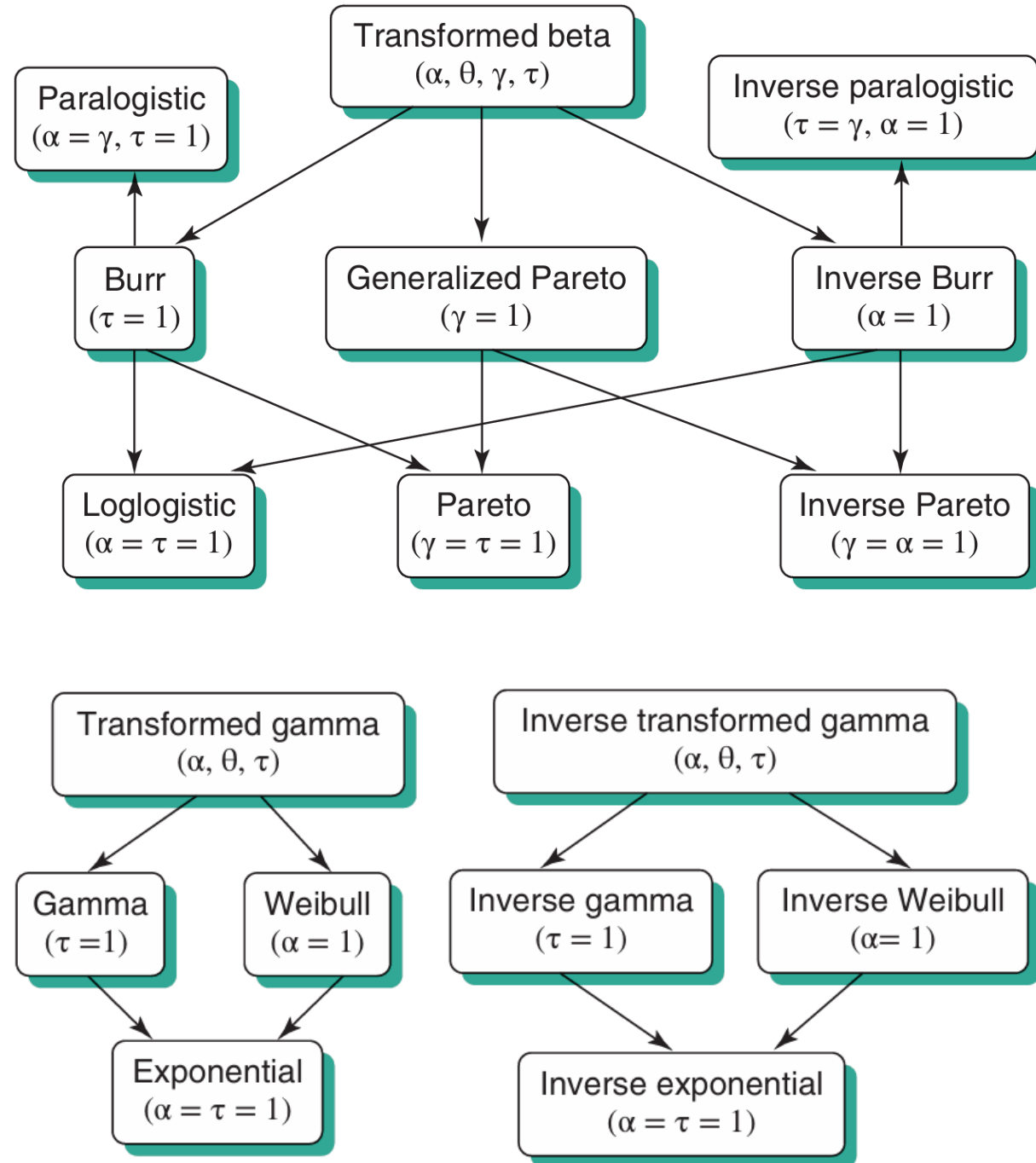
# Selected Distributions and Their Relationships

---

- ❑ Most insurance claims are generally infrequent but can be extremely large, we look for probability distributions "**Loss Models**" that have **positive** support and all tend to be **skewed to the right**.
- ❑ In actuarial science, many named distributions can be related through common construction mechanisms.
- ❑ We need a group of distributions for modeling the "**body**" of a loss distribution.
- ❑ While these distributions can be skewed, they are usually considered "light-tailed" or "medium-tailed" because their tails decay exponentially.
- ❑ Moreover, we need a group of distributions that contains the **heavy-tailed** distributions.
- ❑ These are essential for modeling catastrophic losses where the probability of a very large claim remains significant.

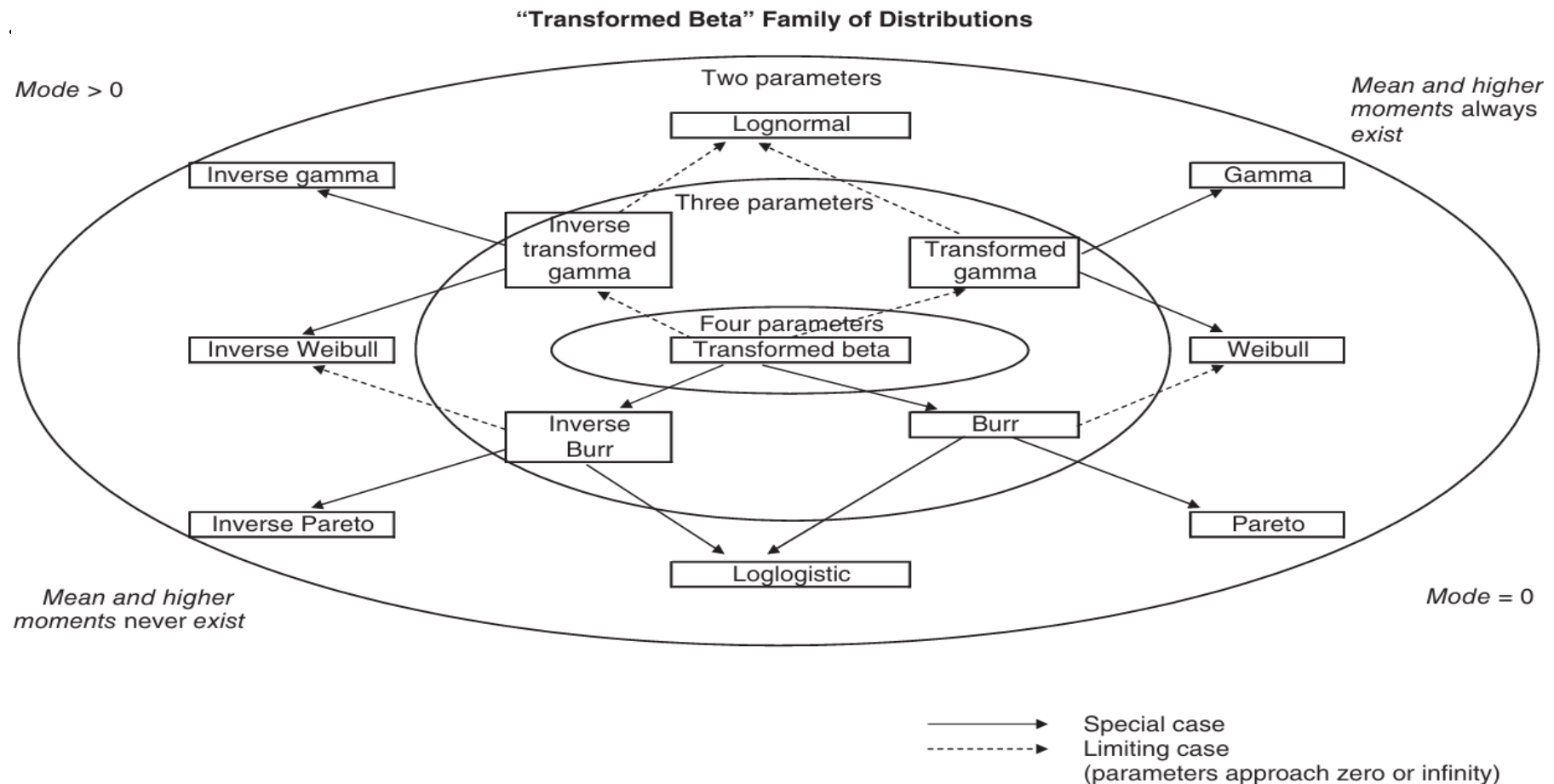
# Two Parametric Families

- Many of distributions “models” can be organized into groupings, that is **distributions are special cases of other distributions**.
- There are two important generator frameworks
  - Beta-generated family
  - Gamma-based family,
- Although many other distributions arise from different transformations or limiting processes.
- For example, the **transformed beta family** includes two special cases of a different nature.
  - Bur  $\rightarrow$  Para-logistic
  - Inverse Bur  $\rightarrow$  Inverse Para-logistic



# Limiting Distributions

- Another way to relate distributions is to see what happens as **parameters** go to their **limiting values of zero or infinity**.
- Mathematically the Transformed Gamma is a limiting case of the Transformed Beta
  - allowing the scale parameter  $\theta \rightarrow \infty$  and the shape parameter  $\alpha \rightarrow \infty$  while maintaining a fixed ratio  $\frac{\theta}{\alpha^{1/\gamma}} \rightarrow \xi$ , the "heavy-tailed" Beta family collapses into the "lighter-tailed" Gamma family.



# The Linear Exponential Family (LEF)

---

□ It is a class of probability distributions whose density PDF (or mass PMF) function can be expressed in a specific mathematical form.

A variable  $X$  belongs to the LEF if its PDF or PMF can be written as:

$$f(x; \theta) = \frac{p(x)e^{r(\theta)x}}{q(\theta)},$$

$p(x)$  depends only on  $x$ , and the function  $q(\theta)$  is a normalizing constant.

## Example

Show that the normal distribution is a member of the linear exponential family.

## Solution

PDF for a Normal distribution:

$$f(x; \mu) = \frac{1}{\sqrt{2\pi\nu}} \exp\left(-\frac{(x-\mu)^2}{2\nu}\right) = \frac{1}{\sqrt{2\pi\nu}} \exp\left(-\frac{x^2 - 2x\mu + \mu^2}{2\nu}\right)$$

split the exponential term into three distinct parts

$$f(x; \mu) = \frac{1}{\sqrt{2\pi\nu}} \cdot \exp\left(-\frac{x^2}{2\nu}\right) \cdot \exp\left(\frac{x\mu}{\nu}\right) \cdot \exp\left(-\frac{\mu^2}{2\nu}\right) = \frac{\left[(2\pi\nu)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2\nu}\right)\right] \cdot \exp\left(\frac{\mu}{\nu}x\right)}{\exp\left(\frac{\mu^2}{2\nu}\right)}.$$

$$\text{Let } p(x) = (2\pi\nu)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2\nu}\right), r(\theta) = \frac{\mu}{\nu} \text{ and } q(\theta) = \exp\left(\frac{\mu^2}{2\nu}\right)$$