

"Other" Distributions: Gamma, Beta, and Dirichlet

Advanced Probability Distributions

DRME

December 22, 2025

Introduction

Why study these distributions?

- **Gamma:** Generalization of Exponential
- **Beta:** Flexible distribution on $[0,1]$ (Frequencies)
- **Dirichlet:** Multivariate generalization of Beta
- Used extensively in:
 - Bayesian statistics (Evolutionary biology)
 - Machine learning
 - Reliability theory
 - Natural language processing

Connections:

- Exponential \subset Gamma \subset Chi-square
- Beta is conjugate prior for Bernoulli/Binomial
- Dirichlet is conjugate prior for Multinomial

Gamma Distribution

Gamma Distribution: Definition

Definition (Gamma Distribution)

A continuous random variable X has a **Gamma** distribution with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$ if its PDF is:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0$$

where $\Gamma(\alpha)$ is the gamma function:

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

Notation: $X \sim \text{Gamma}(\alpha, \beta)$

Alternative parameterization: Sometimes uses scale parameter $\theta = 1/\beta$:

$$f_X(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$$

[Wikipedia: Gamma Distribution](#)

Gamma Function

The **gamma function** $\Gamma(\alpha)$ is crucial for the Gamma distribution:

Properties:

1. $\Gamma(1) = 1$
2. $\Gamma(1/2) = \sqrt{\pi}$
3. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for $\alpha > 0$
4. For integer n : $\Gamma(n) = (n - 1)!$
5. $\Gamma(\alpha)\Gamma(1 - \alpha) = \frac{\pi}{\sin(\pi\alpha)}$

Connection to factorial: For positive integers n :

$$\Gamma(n) = (n - 1)! \quad \text{so} \quad \Gamma(n + 1) = n!$$

Why needed?: Ensures PDF integrates to 1:

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

Special Cases of Gamma Distribution

The Gamma distribution includes several important special cases:

1. **Exponential(λ):** Gamma($1, \lambda$)

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0$$

2. **Chi-square with k degrees of freedom:** Gamma($k/2, 1/2$)

$$f_X(x) = \frac{1}{2^{k/2} \Gamma(k/2)} x^{k/2-1} e^{-x/2}, \quad x > 0$$

3. **Erlang(n, λ):** Gamma(n, λ) where n is integer
 - Sum of n i.i.d. Exponential(λ) random variables
 - Models waiting time for n events in Poisson process

Moments of Gamma Distribution

Mean and Variance:

$$E[X] = \frac{\alpha}{\beta}, \quad \text{Var}(X) = \frac{\alpha}{\beta^2}$$

Moment Generating Function:

$$M_X(t) = E[e^{tX}] = \left(1 - \frac{t}{\beta}\right)^{-\alpha}, \quad t < \beta$$

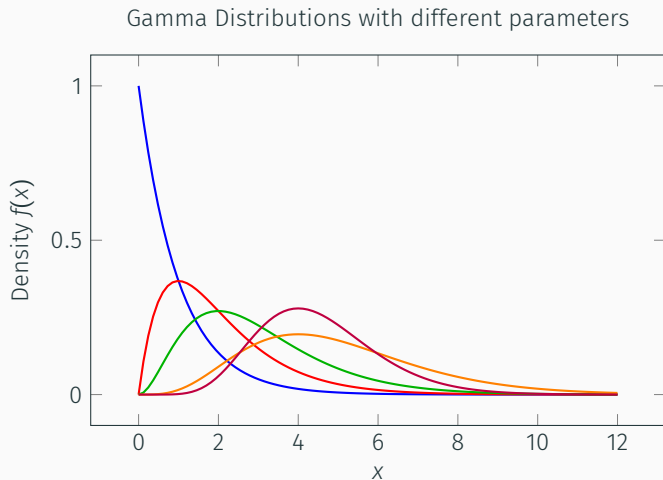
Skewness and Kurtosis:

$$\text{Skewness} = \frac{2}{\sqrt{\alpha}}, \quad \text{Excess Kurtosis} = \frac{6}{\alpha}$$

Interpretation:

- α controls shape: small α = more skewed, large $\alpha \approx$ Normal
- β controls scale: larger β = more concentrated near 0

Visualizing Gamma Distribution



— Gamma(1, 1) (Exponential)

— Gamma(2, 1)

— Gamma(3, 1)

— Gamma(5, 1)

— Gamma(2, 2)

Additivity Property

Theorem (Additivity of Gamma)

If X_1, \dots, X_n are independent with $X_i \sim \text{Gamma}(\alpha_i, \beta)$, then:

$$\sum_{i=1}^n X_i \sim \text{Gamma}\left(\sum_{i=1}^n \alpha_i, \beta\right)$$

Proof: Use MGFs:

$$M_{\sum X_i}(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n \left(1 - \frac{t}{\beta}\right)^{-\alpha_i} = \left(1 - \frac{t}{\beta}\right)^{-\sum \alpha_i}$$

Important Consequences:

1. Sum of i.i.d. Exponential(β) \sim Gamma(n, β)
2. Sum of independent Chi-squares \sim Chi-square (sum of df)
3. Waiting time for n events in Poisson(β) process \sim Gamma(n, β)

Memoryless Property (Lack Thereof)

Important Distinction: Unlike Exponential, Gamma is **not** memoryless for $\alpha > 1$.

Why?: For $X \sim \text{Gamma}(\alpha, \beta)$ with $\alpha > 1$:

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

Intuition:

- Exponential: Constant hazard rate
- Gamma with $\alpha > 1$: Increasing hazard rate (wear-out)
- Gamma with $\alpha < 1$: Decreasing hazard rate (burn-in)

Hazard function:

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha, \beta x)}$$

where $\Gamma(\alpha, z) = \int_z^\infty t^{\alpha-1} e^{-t} dt$ is the incomplete gamma function.

Application: Reliability Engineering

Problem: A system consists of n identical components with Exponential(λ) lifetimes. System fails when all components fail (series system). What's the system lifetime distribution?

Solution: System lifetime = minimum of component lifetimes

$$X_{\text{system}} = \min(X_1, \dots, X_n), \quad X_i \sim \text{Exp}(\lambda)$$

For series system:

$$P(X_{\text{system}} > t) = P(\min_i X_i > t) = \prod_{i=1}^n P(X_i > t) = (e^{-\lambda t})^n = e^{-n\lambda t}$$

So $X_{\text{system}} \sim \text{Exp}(n\lambda)$

Parallel system: System fails when all components fail

$$X_{\text{system}} = \max(X_1, \dots, X_n)$$

Distribution is **not** Gamma but can be derived from order statistics.

Application: Bayesian Statistics

Gamma as conjugate prior for Poisson rate parameter:

Setting:

- Likelihood: $X_1, \dots, X_n \sim \text{Pois}(\lambda)$ i.i.d.
- Prior: $\lambda \sim \text{Gamma}(\alpha_0, \beta_0)$

Posterior distribution:

$$\begin{aligned} p(\lambda \mid \text{data}) &\propto \left[\prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \right] \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda^{\alpha_0-1} e^{-\beta_0 \lambda} \\ &\propto \lambda^{\alpha_0 + \sum x_i - 1} e^{-(\beta_0 + n)\lambda} \end{aligned}$$

So:

$$\lambda \mid \text{data} \sim \text{Gamma} \left(\alpha_0 + \sum_{i=1}^n x_i, \beta_0 + n \right)$$

Interpretation: Prior pseudo-counts α_0 , pseudo-time β_0 updated by actual data.

Beta Distribution

Beta Distribution: Definition

Definition (Beta Distribution)

A continuous random variable X has a **Beta** distribution with parameters $\alpha > 0$ and $\beta > 0$ if its PDF is:

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

where $B(\alpha, \beta)$ is the beta function:

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Notation: $X \sim \text{Beta}(\alpha, \beta)$

Support: $(0, 1)$ - naturally models probabilities, proportions, percentages

[Wikipedia: Beta Distribution](#)

Beta Function and Its Properties

The **beta function** $B(\alpha, \beta)$ normalizes the Beta PDF:

Properties:

1. Symmetry: $B(\alpha, \beta) = B(\beta, \alpha)$
2. Relation to Gamma: $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
3. Special values:
 - $B(1, 1) = 1$
 - $B(1/2, 1/2) = \pi$
 - $B(m, n) = \frac{(m-1)!(n-1)!}{(m+n-1)!}$ for integers m, n
4. Recursion: $B(\alpha + 1, \beta) = \frac{\alpha}{\alpha + \beta} B(\alpha, \beta)$

Incomplete Beta Function:

$$B(x; \alpha, \beta) = \int_0^x t^{\alpha-1} (1-t)^{\beta-1} dt, \quad 0 \leq x \leq 1$$

CDF of Beta distribution: $F_X(x) = \frac{B(x; \alpha, \beta)}{B(\alpha, \beta)}$

Moments of Beta Distribution

Mean and Variance:

$$E[X] = \frac{\alpha}{\alpha + \beta}, \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Higher Moments:

$$E[X^k] = \frac{B(\alpha + k, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha + k)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\alpha + \beta + k)}$$

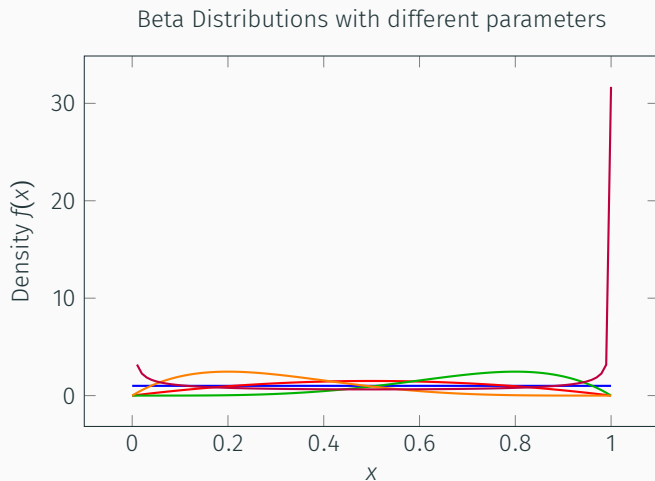
Mode (for $\alpha, \beta > 1$):

$$\text{mode} = \frac{\alpha - 1}{\alpha + \beta - 2}$$

Interpretation of parameters:

- α : "number of successes" + 1
- β : "number of failures" + 1
- $\alpha + \beta$: "total sample size" + 2

Visualizing Beta Distribution



- Beta(1, 1) (Uniform)
- Beta(2, 2)
- Beta(5, 2)
- Beta(2, 5)

Special Cases:

1. $\text{Beta}(1, 1) = \text{Uniform}(0, 1)$
2. $\text{Beta}(1/2, 1/2)$: Arcsine distribution
3. $\text{Beta}(\alpha, 1)$: Power function distribution

Relationships with other distributions:

1. If $X \sim \text{Beta}(\alpha, \beta)$, then $1 - X \sim \text{Beta}(\beta, \alpha)$
2. If $U_1, \dots, U_{\alpha+\beta-1} \sim \text{Uniform}(0, 1)$ i.i.d., then $U_{(\alpha)} \sim \text{Beta}(\alpha, \beta)$
3. If $X \sim \text{Gamma}(\alpha, 1)$ and $Y \sim \text{Gamma}(\beta, 1)$ independent, then:

$$\frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta)$$

Order Statistics Connection

Fundamental relationship: Order statistics of Uniform distribution follow Beta distribution.

Theorem: Let $U_1, \dots, U_n \sim \text{Uniform}(0, 1)$ i.i.d. Then the k -th order statistic $U_{(k)}$ has distribution:

$$U_{(k)} \sim \text{Beta}(k, n - k + 1)$$

Proof intuition:

- Need $k - 1$ variables below u , 1 variable at u , $n - k$ variables above u
- Each arrangement has probability $u^{k-1} \cdot du \cdot (1 - u)^{n-k}$
- Number of arrangements: $\binom{n}{k-1, 1, n-k} = \frac{n!}{(k-1)!1!(n-k)!}$
- After normalization: PDF matches $\text{Beta}(k, n - k + 1)$

Application: Distribution of sample quantiles, nonparametric statistics.

Application: Bayesian Statistics (Binomial)

Beta as conjugate prior for Binomial success probability:

Setting:

- Likelihood: $X \sim \text{Bin}(n, p)$
- Prior: $p \sim \text{Beta}(\alpha, \beta)$

Posterior distribution:

$$\begin{aligned} p(p \mid X = x) &\propto \binom{n}{x} p^x (1-p)^{n-x} \times \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} \\ &\propto p^{\alpha+x-1} (1-p)^{\beta+n-x-1} \end{aligned}$$

So:

$$p \mid X = x \sim \text{Beta}(\alpha + x, \beta + n - x)$$

Interpretation: Prior counts $\alpha - 1$ successes, $\beta - 1$ failures updated by x successes, $n - x$ failures.

Application: Project Management (PERT)

Problem: Estimate task duration with optimistic (a), most likely (m), and pessimistic (b) estimates.

PERT Beta Distribution: Assume duration $X \sim \text{Beta}(\alpha, \beta)$ scaled to $[a, b]$:

$$X = a + (b - a)Y, \quad Y \sim \text{Beta}(\alpha, \beta)$$

Parameter estimation:

$$\text{Mean} = E[X] = a + (b - a) \frac{\alpha}{\alpha + \beta}$$

$$\text{Mode} = m = a + (b - a) \frac{\alpha - 1}{\alpha + \beta - 2}$$

Common approximation: Assume $\text{Beta}(4, 4)$ for Y , then:

$$E[X] = \frac{a + 4m + b}{6}, \quad \text{Var}(X) = \left(\frac{b - a}{6} \right)^2$$

This is the classic PERT formula.

Dirichlet Distribution

Dirichlet Distribution: Definition

Definition (Dirichlet Distribution)

A K -dimensional random vector $\mathbf{X} = (X_1, \dots, X_K)$ has a **Dirichlet** distribution with parameters $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$, $\alpha_i > 0$, if its PDF is:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^K x_i^{\alpha_i-1}, \quad \mathbf{x} \in S_K$$

where S_K is the $(K-1)$ -dimensional simplex:

$$S_K = \left\{ (x_1, \dots, x_K) : x_i \geq 0, \sum_{i=1}^K x_i = 1 \right\}$$

and $B(\boldsymbol{\alpha})$ is the multivariate beta function:

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\alpha_0)}, \quad \alpha_0 = \sum_{i=1}^K \alpha_i$$

Notation: $\mathbf{X} \sim \text{Dir}(\boldsymbol{\alpha})$

Understanding the Dirichlet Distribution

Key Properties:

- **Support:** Probability simplex S_K (vectors of probabilities summing to 1)
- **Dimension:** Only $K - 1$ free dimensions due to sum constraint
- **Interpretation:** Distribution over probability vectors
- **Parameters:** $\alpha_i > 0$ control concentration in each component

Marginal Distributions:

$$X_i \sim \text{Beta}(\alpha_i, \alpha_0 - \alpha_i)$$

where $\alpha_0 = \sum_{j=1}^K \alpha_j$

Aggregation Property:

$$(X_i + X_j, X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_K) \sim \text{Dir}(\alpha_i + \alpha_j, \alpha_1, \dots)$$

[Wikipedia: Dirichlet Distribution](#)

Moments of Dirichlet Distribution

Means:

$$E[X_i] = \frac{\alpha_i}{\alpha_0}, \quad \alpha_0 = \sum_{j=1}^K \alpha_j$$

Variances and Covariances:

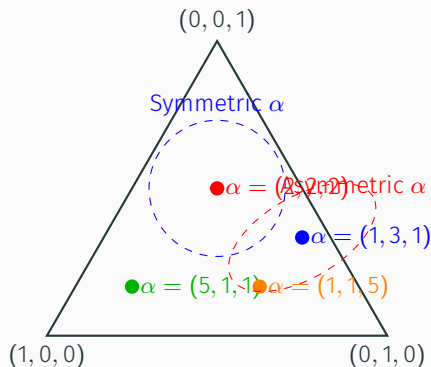
$$\text{Var}(X_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}$$

$$\text{Cov}(X_i, X_j) = -\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)} \quad (i \neq j)$$

Interpretation:

- α_i/α_0 : Expected proportion in category i
- α_0 : Concentration parameter
 - Small α_0 : Sparse distributions (few dominant categories)
 - Large α_0 : Uniform-like distributions

Visualizing Dirichlet Distribution (K=3)



Observations:

- Points in triangle represent (x_1, x_2, x_3) with $x_1 + x_2 + x_3 = 1$
- Symmetric α : Concentration near center
- Asymmetric α : Concentration near vertices
- Small α_0 : More spread out
- Large α_0 : Tight concentration around mean

Dirichlet as Multivariate Beta

Relationship to Beta: Dirichlet is multivariate generalization of Beta:

- Beta: Distribution on 1D simplex (interval $[0,1]$)
- Dirichlet: Distribution on $(K - 1)$ -D simplex

Construction from Gamma variables: If $Y_1, \dots, Y_K \sim \text{Gamma}(\alpha_i, 1)$ independent, then:

$$\left(\frac{Y_1}{\sum_{j=1}^K Y_j}, \dots, \frac{Y_K}{\sum_{j=1}^K Y_j} \right) \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$$

Uniform Dirichlet: $\text{Dir}(1, \dots, 1)$ is uniform distribution over simplex

Symmetric Dirichlet: $\text{Dir}(\alpha, \dots, \alpha)$

- All categories equally likely on average
- α controls concentration

Application: Bayesian Statistics (Multinomial)

Dirichlet as conjugate prior for Multinomial parameters:

Setting:

- Likelihood: $\mathbf{X} = (X_1, \dots, X_K) \sim \text{Multinomial}(n, \mathbf{p})$
- Prior: $\mathbf{p} \sim \text{Dir}(\boldsymbol{\alpha})$

Posterior distribution:

$$\begin{aligned} p(\mathbf{p} \mid \mathbf{X} = \mathbf{x}) &\propto \left[\prod_{i=1}^K p_i^{x_i} \right] \times \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^K p_i^{\alpha_i - 1} \\ &\propto \prod_{i=1}^K p_i^{\alpha_i + x_i - 1} \end{aligned}$$

So:

$$\mathbf{p} \mid \mathbf{X} = \mathbf{x} \sim \text{Dir}(\alpha_1 + x_1, \dots, \alpha_K + x_K)$$

Interpretation: Prior pseudo-counts $\alpha_i - 1$ updated by observed counts x_i .

Application: Natural Language Processing (LDA)

Latent Dirichlet Allocation (LDA): Generative model for documents

Model structure:

1. For each topic k :

$$\phi_k \sim \text{Dir}(\beta) \quad (\text{word distribution})$$

2. For each document d :

$$\theta_d \sim \text{Dir}(\alpha) \quad (\text{topic distribution})$$

3. For each word n in document d :

- Choose topic: $z_{dn} \sim \text{Multinomial}(\theta_d)$
- Choose word: $w_{dn} \sim \text{Multinomial}(\phi_{z_{dn}})$

Key insight: Dirichlet priors provide natural way to model distributions over:

- Topics in documents (θ_d)
- Words in topics (ϕ_k)

Application: Machine Learning (Dirichlet Process)

Dirichlet Process: Nonparametric Bayesian method

Idea: Distribution over distributions

$$G \sim DP(\alpha, H)$$

- α : Concentration parameter
- H : Base distribution

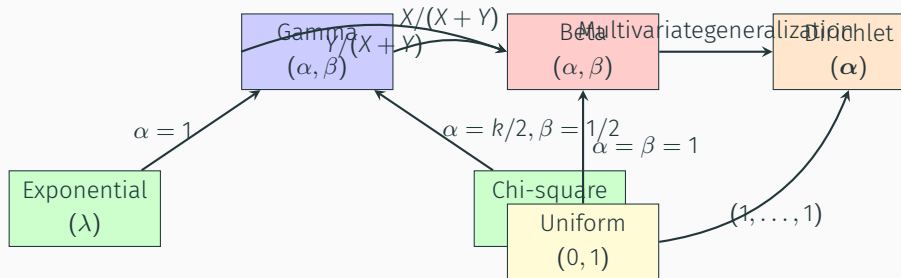
Chinese Restaurant Process (CRP) representation:

- Customers (data points) enter restaurant
- Customer $n + 1$ sits at:
 - Existing table k with probability $\frac{n_k}{\alpha + n}$
 - New table with probability $\frac{\alpha}{\alpha + n}$
- n_k : number of customers at table k

Applications: Clustering, topic modeling, mixture models with unknown number of components.

Interconnections and Summary

Relationships Between Distributions



Key relationships:

- Exponential, Chi-square \subset Gamma
- Beta = ratio of independent Gammas
- Dirichlet = multivariate Beta
- Uniform = special case of Beta and Dirichlet

Conjugate Prior Summary

Bayesian inference: Prior and posterior in same distribution family

Likelihood	Parameter	Conjugate Prior	Posterior
Bernoulli(p)	p	Beta(α, β)	Beta($\alpha + x, \beta + n - x$)
Binomial(n, p)	p	Beta(α, β)	Beta($\alpha + x, \beta + n - x$)
Poisson(λ)	λ	Gamma(α, β)	Gamma($\alpha + \sum x_i, \beta + n$)
Multinomial(n, \mathbf{p})	\mathbf{p}	Dirichlet($\boldsymbol{\alpha}$)	Dirichlet($\boldsymbol{\alpha} + \mathbf{x}$)
Normal(μ, σ^2) known σ^2	μ	Normal(μ_0, σ_0^2)	Normal(μ_n, σ_n^2)

Why conjugate priors?

- Computational convenience
- Interpretable parameters (pseudo-counts)
- Easy updating with new data

Key Formulas and Properties

Gamma (α, β):

- PDF: $\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
- Mean: α/β
- Variance: α/β^2
- Additive:
$$\sum \text{Gamma}(\alpha_i, \beta) = \text{Gamma}(\sum \alpha_i, \beta)$$

Beta (α, β):

- PDF: $\frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$
- Mean: $\alpha/(\alpha + \beta)$
- Variance: $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Marginal of Dirichlet

Dirichlet (α):

- PDF: $\frac{1}{B(\alpha)} \prod x_i^{\alpha_i-1}$
- Mean: α_i/α_0
- $\alpha_0 = \sum \alpha_i$
- Covariance: negative

Common Applications:

- Gamma: Waiting times, reliability, Bayesian rates
- Beta: Proportions, probabilities, Bayesian success rates
- Dirichlet: Multinomial probabilities, topic models, mixture models

When to use each distribution?

Gamma Distribution

- Modeling waiting times for multiple events
- Sum of Exponential random variables
- Bayesian inference for Poisson rate
- Prior for precision in Normal distribution

Beta Distribution

- Modeling probabilities/proportions
- Bayesian inference for Binomial p
- Order statistics of Uniform
- Flexible distribution on $[0,1]$

Dirichlet Distribution

- Modeling probability vectors (sum to 1)
- Bayesian inference for Multinomial

Resources and Further Reading

Comprehensive Wikipedia articles for further study:

Gamma Distribution



Main Article:
[Gamma Distribution](#)

Related Articles:

- [Gamma Function](#)
- [Exponential Distribution](#)
- [Chi-square Distribution](#)
- [Erlang Distribution](#)

Beta Distribution

Main Article:
[Beta Distribution](#)

Related Articles:

- [Beta Function](#)
- [Uniform Distribution](#)
- [Order Statistics](#)
- [Conjugate Prior](#)

Dirichlet Distribution

Main Article:
[Dirichlet Distribution](#)

Related Articles:

- [Dirichlet Process](#)
- [Latent Dirichlet Allocation](#)
- [Multinomial Distribution](#)
- [Probability Simplex](#)

Interactive Visualizations and Tutorials:

- **Distributions.jl**: Interactive visualizations of probability distributions
<https://juliastats.org/Distributions.jl/stable/>
- **Stat Trek Tutorial**: Gamma and Beta distributions explained
<https://stattrek.com/probability-distributions/gamma>
- **Probability Course**: Online textbook with examples
https://www.probabilitycourse.com/chapter4/4_2_4_Gamma_distribution.php
- **3Blue1Brown**: Visual explanations of probability concepts
<https://www.3blue1brown.com/topics/probability>

Software Implementations:

- R: `dgamma()`, `dbeta()`, `ddirichlet()` functions
- Python: `scipy.stats.gamma`, `scipy.stats.beta`,
`numpy.random.dirichlet`
- Julia: `Distributions.Gamma()`, `Distributions.Beta()`,
`Distributions.Dirichlet()`

Historical Development:

- **Gamma Function:** First studied by Euler (1729), extended by Legendre, Gauss
- **Gamma Distribution:** Introduced by Laplace (1836), further developed by Pearson (1895)
- **Beta Distribution:** Studied by Thomas Bayes (posthumous 1763), Karl Pearson (1895)
- **Dirichlet Distribution:** Named after Peter Gustav Lejeune Dirichlet (1805-1859)

- **Bayesian Statistics:** Conjugate priors
- **Machine Learning:** Topic modeling (LDA)
- **Reliability Engineering:** Failure time analysis
- **Bioinformatics:** Gene expression analysis / Phylogeny
- **Economics:** Risk modeling
- **Physics:** Particle physics simulations
- **Psychology:** Psychometric modeling
- **Ecology:** Species distribution modeling

Quick Reference: Key Relationships

Relationship	Mathematical Expression
Gamma \rightarrow Exponential	$\text{Gamma}(1, \lambda) = \text{Exp}(\lambda)$
Gamma \rightarrow Chi-square	$\text{Gamma}(k/2, 1/2) = \chi^2(k)$
Gamma \rightarrow Beta	$\frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta)$ where $X \sim \text{Gamma}(\alpha, 1)$, $Y \sim \text{Gamma}(\beta, 1)$
Beta \rightarrow Uniform	$\text{Beta}(1, 1) = \text{Uniform}(0, 1)$
Beta \rightarrow Dirichlet (marginal)	$X_i \sim \text{Beta}\left(\alpha_i, \sum_{j \neq i} \alpha_j\right)$ where $(X_1, \dots, X_K) \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$
Order Statistics	$U_{(k)} \sim \text{Beta}(k, n - k + 1)$ where $U_1, \dots, U_n \sim \text{Uniform}(0, 1)$ i.i.d.
Dirichlet from Gamma	$\left(\frac{Y_1}{\sum Y_j}, \dots, \frac{Y_K}{\sum Y_j}\right) \sim \text{Dir}(\alpha_1, \dots, \alpha_K)$ where $Y_i \sim \text{Gamma}(\alpha_i, 1)$ independent

Basic Problems:

1. If $X \sim \text{Gamma}(3, 2)$, find $E[X]$ and $\text{Var}(X)$.
2. Show that if $X \sim \text{Beta}(\alpha, \beta)$, then $1 - X \sim \text{Beta}(\beta, \alpha)$.
3. If $(X_1, X_2, X_3) \sim \text{Dir}(2, 3, 5)$, find $E[X_1]$ and $\text{Cov}(X_1, X_2)$.
4. Prove that the sum of n i.i.d. $\text{Exp}(\lambda)$ random variables follows $\text{Gamma}(n, \lambda)$.

Intermediate Problems:

5. Show that the mode of $\text{Beta}(\alpha, \beta)$ is $(\alpha - 1)/(\alpha + \beta - 2)$ for $\alpha, \beta > 1$.
6. If $X \sim \text{Gamma}(\alpha, 1)$ and $Y \sim \text{Gamma}(\beta, 1)$ are independent, show that $X/(X + Y) \sim \text{Beta}(\alpha, \beta)$.
7. Derive the marginal distribution of X_i from $\text{Dir}(\alpha_1, \dots, \alpha_K)$.

Advanced Problems:

8. Show that $\text{Dir}(1, \dots, 1)$ is the uniform distribution over the probability simplex.
9. Prove the aggregation property of the Dirichlet distribution.
10. Derive the posterior distribution when using a $\text{Dir}(\boldsymbol{\alpha})$ prior for a $\text{Multinomial}(n, \mathbf{p})$ likelihood.