Review: Random Variables and Probability Distributions

Discrete and Continuous Distributions — Definitions, Properties, and Examples

Prof. Michael Mascagni

Department of Computer Science
Department of Mathematics
Department of Scientific Computing
Graduate Program in Molecular Biophysics
Florida State University, Tallahassee, FL 32306 **USA**
E-mail: mascagni@fsu.edu
URL: http://www.cs.fsu.edu/∼mascagni

Outline

Random Variables

Definition

A **random variable** X is a function $X : \Omega \to \mathbb{R}$ that maps each outcome of a sample space Ω to a real number.

Two fundamental types:

- **Discrete:** Takes a countable set of values $\{x_1, x_2, x_3, \ldots\}$.
  Example: number of heads in 10 coin flips $\in \{0, 1, \ldots, 10\}$.
- **Continuous:** Takes values in an interval (or union of intervals) in $\mathbb{R}$.
  Example: time (in seconds) until a server responds $\in [0, \infty)$.

Notation

- X — the random variable (capital letter, before observation)
- x — a specific realized value (lowercase)
- $P(X = x)$ — probability that X takes the value x
- $P(a \leq X \leq b)$ — probability that X falls in $[a, b]$

Probability Mass Function (PMF)

Definition — Discrete Random Variables

The **probability mass function** (PMF) of a discrete random variable X assigns a probability to every possible value:

$$p(x) \;=\; P(X = x), \qquad x \in \mathbb{R}$$

Required properties:

1. $p(x) \geq 0$ for all x
2. $\displaystyle\sum_{\text{all } x} p(x) = 1$
3. $\displaystyle P(a \leq X \leq b) = \sum_{a \leq x \leq b} p(x)$

Example: Fair six-sided die: $p(x) = \frac{1}{6}$ for $x \in \{1, 2, 3, 4, 5, 6\}$, $\quad p(x) = 0$ otherwise.

Alternate name: Also called the **probability function** or **discrete density function**.

Probability Density Function (PDF)

Definition — Continuous Random Variables

A continuous random variable X has a **probability density function** (PDF) $f(x)$ such that for any interval $[a, b]$:

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

Required properties:

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$

Key Distinction: Density $\neq$ Probability

For a continuous RV, $P(X = x) = 0$ for every single point x. Probabilities require integration over an *interval*. $f(x)$ can exceed 1 — it is a density, not a probability.

Example: $f(x) = 2x$ for $x \in [0, 1]$ is a valid PDF: $\int_0^1 2x\, dx = 1$.

Cumulative Distribution Function (CDF)

Definition

The **cumulative distribution function** (CDF) of any random variable X is:

$$F(x) = P(X \leq x), \qquad x \in \mathbb{R}$$

Universal properties (discrete and continuous):

1. $0 \leq F(x) \leq 1$ for all x
2. F is non-decreasing: $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$
4. F is right-continuous: $\lim_{h \to 0^+} F(x+h) = F(x)$

Connection to PMF/PDF:

$$\text{Discrete: } F(x) = \sum_{t \leq x} p(t) \qquad \text{Continuous: } F(x) = \int_{-\infty}^{x} f(t)\,dt \iff f(x) = F'(x)$$

Useful identity: $P(a < X \leq b) = F(b) - F(a)$

Expected Value and Mean

Definition

The **expected value** (or **mean**) of X, written $E[X]$ or μ, is the probability-weighted average of all outcomes.

$$\mu = E[X] = \begin{cases} \sum_{x} x\, p(x) & \text{discrete} \\ \int_{-\infty}^{\infty} x\, f(x)\, dx & \text{continuous} \end{cases}$$

Key properties:

- **Linearity:** $E[aX + b] = a\,E[X] + b$
- **Additivity:** $E[X + Y] = E[X] + E[Y]$ (always, no independence needed)
- **LOTUS** (Law of the Unconscious Statistician): $E[g(X)] = \sum_x g(x)\, p(x)$ or $E[g(X)] = \int g(x)\, f(x)\, dx$

Variance and Standard Deviation

Definition

The **variance** of X measures the expected squared deviation from the mean:

$$\sigma^2 = \mathrm{Var}(X) = E\left[(X-\mu)^2\right]$$

Computational shortcut:

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

Standard deviation:

$$\sigma = \sqrt{\mathrm{Var}(X)} \qquad \text{(same units as } X\text{)}$$

Key properties:

- $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$
- $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ if X, Y are *independent*
- $\mathrm{Var}(X) \geq 0$ always; $\mathrm{Var}(X) = 0$ iff X is a constant

Moment Generating Function (MGF)

Definition

The **moment generating function** (MGF) of X is (Laplace transform):

$$M_X(t) \;=\; E\left[e^{tX}\right], \qquad t \in (-h, h) \text{ for some } h > 0$$

Why "moment generating"? Expanding e^{tX} in a power series:

$$M_X(t) = E\left[\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{E[X^n]}{n!}\, t^n$$

So the n-th derivative at $t = 0$ gives the n-th **moment**:

$$M_X^{(n)}(0) = E[X^n] \qquad \Longrightarrow \qquad \mu = M_X'(0), \quad \sigma^2 = M_X''(0) - \left[M_X'(0)\right]^2$$

Uniqueness Theorem

If $M_X(t) = M_Y(t)$ for all t in some open interval containing 0, then X and Y have the *same* probability distribution. MGFs uniquely identify distributions.

Bernoulli Distribution

Setup: A single trial with two outcomes — success (prob. p) or failure (prob. $1-p$).

Notation: $X \sim \mathrm{Bernoulli}(p), \quad 0 < p < 1$

PMF:

$$p(x) = p^x(1-p)^{1-x}, \quad x \in \{0, 1\}$$

CDF:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

Mean:

$$\mu = p$$

Variance:

$$\sigma^2 = p(1-p)$$

MGF:

$$M(t) = 1 - p + pe^t$$

valid for all $t \in \mathbb{R}$

Example: Single coin flip with $p = 0.5$: $\mu = 0.5$, $\sigma^2 = 0.25$ (maximum possible variance).

Note: The Bernoulli is the building block for Binomial, Geometric, and Negative Binomial distributions.

Binomial Distribution

Setup: Count of successes in n *independent* Bernoulli(p) trials.

Notation: $X \sim \text{Binomial}(n, p)$

PMF:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

for $x = 0, 1, \ldots, n$

CDF:

$$F(k) = \sum_{x=0}^{k} \binom{n}{x} p^x (1-p)^{n-x}$$

No closed form; tabulated or computed via the regularized incomplete beta function $I_{1-p}(n-k,\ k+1)$.

Mean:

$$\mu = np$$

Variance:

$$\sigma^2 = np(1-p)$$

MGF:

$$M(t) = \left(1 - p + pe^t\right)^n$$

valid for all $t \in \mathbb{R}$

Example: 20 exam questions, each guessed with $p = 0.25$: $\mu = 5$ correct, $\sigma^2 = 3.75$.

Geometric Distribution

Setup: Number of independent Bernoulli(p) trials needed to obtain the *first* success.

Notation: $X \sim \mathrm{Geometric}(p)$

PMF:

$$p(x) = (1-p)^{x-1}\, p, \quad x = 1, 2, 3, \ldots$$

CDF:

$$F(k) = 1 - (1-p)^{k}, \quad k = 1, 2, 3, \ldots$$

Mean:

$$\mu = \frac{1}{p}$$

Variance:

$$\sigma^2 = \frac{1-p}{p^2}$$

MGF:

$$M(t) = \frac{p\, e^t}{1-(1-p)e^t}$$

valid for $t < -\ln(1-p)$

Memoryless Property (Discrete)

$P(X > m+n \mid X > m) = P(X > n)$ for all $m, n > 0$. The Geometric is the *only* discrete memoryless

Negative Binomial Distribution

Setup: Number of independent Bernoulli(p) trials needed to obtain the r-th success.

Notation: $X \sim \mathrm{NegBin}(r,\, p)$

PMF:

$$p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

for $x = r,\ r+1,\ r+2, \ldots$

CDF:

$$F(k) = I_p(r,\, k-r+1)$$

where I_p is the regularized incomplete beta function.

Mean:

$$\mu = \frac{r}{p}$$

Variance:

$$\sigma^2 = \frac{r(1-p)}{p^2}$$

MGF:

$$M(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$$

valid for $t < -\ln(1-p)$

Relationship: $r = 1$ recovers Geometric(p). The NegBin(r, p) equals the sum of r independent Geometric(p) variables.

Poisson Distribution

Setup: Number of events occurring in a fixed interval when events arrive at constant mean rate λ and independently.

Notation: $X \sim \mathrm{Poisson}(\lambda), \quad \lambda > 0$

PMF:

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

CDF:

$$F(k) = e^{-\lambda}\sum_{x=0}^{k}\frac{\lambda^x}{x!} = \frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor!}$$

(regularized upper incomplete gamma)

Mean:

$$\mu = \lambda$$

Variance:

$$\sigma^2 = \lambda$$

MGF:

$$M(t) = e^{\lambda(e^t - 1)}$$

valid for all $t \in \mathbb{R}$

Example: A router receives 100 packets/sec. In 1 ms, $\lambda = 0.1$. $\quad P(X = 0) = e^{-0.1} \approx 0.905$.

Limit: As $n \to \infty$, $p \to 0$, $np \to \lambda$: $\mathrm{Binomial}(n, p) \to \mathrm{Poisson}(\lambda)$.

Hypergeometric Distribution

Setup: Drawing n items *without replacement* from a population of N items containing K successes.

Notation: $X \sim \text{Hypergeometric}(N, K, n)$

PMF:

$$p(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

for $\max(0, n+K-N) \leq x \leq \min(n, K)$

CDF: No simple closed form; computed via sum of PMF values.

Mean:

$$\mu = n\,\frac{K}{N}$$

Variance:

$$\sigma^2 = n\frac{K}{N}\,\frac{N-K}{N}\,\frac{N-n}{N-1}$$

Finite population correction:

$$\frac{N-n}{N-1} < 1$$

reduces variance vs. Binomial.

Example: 52-card deck, draw 5. $K = 13$ spades, $n = 5$: $\mu = 5 \cdot 13/52 = 1.25$ spades expected.

Relationship: As $N \to \infty$ with $K/N \to p$: $\text{Hypergeometric}(N, K, n) \to \text{Binomial}(n, p)$.

Continuous Uniform Distribution

Setup: All values in the interval $[a, b]$ are equally likely.

Notation: $X \sim \mathrm{Uniform}(a,\, b)$

PDF:

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$

CDF:

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Mean:

$$\mu = \frac{a+b}{2}$$

Variance:

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

MGF:

$$M(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}, \quad t \neq 0$$

$M(0) = 1$

Application: Foundation of random number generation. If $U \sim \mathrm{Uniform}(0, 1)$ and F is any CDF, then $X = F^{-1}(U)$ has distribution F (Inverse CDF method).

Exponential Distribution

Setup: Time between consecutive events in a Poisson process with rate λ.

Notation: $X \sim \mathrm{Exp}(\lambda), \quad \lambda > 0$

PDF:

$$f(x) = \lambda\, e^{-\lambda x}, \quad x \geq 0$$

CDF:

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

Mean:

$$\mu = \frac{1}{\lambda}$$

Variance:

$$\sigma^2 = \frac{1}{\lambda^2}$$

MGF:

$$M(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda$$

Memoryless Property (Continuous)

$P(X > s + t \mid X > s) = P(X > t)$ for all $s, t \geq 0$. The Exponential is the *only* continuous memoryless distribution.

Normal (Gaussian) Distribution

Setup: The most important distribution in statistics; arises naturally via the Central Limit Theorem.

Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$

PDF:

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

for $x \in (-\infty, \infty)$

CDF:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

Mean:

$$E[X] = \mu$$

Variance:

$$\mathrm{Var}(X) = \sigma^2$$

MGF:

$$M(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

valid for all $t \in \mathbb{R}$

Standardization: $Z = (X-\mu)/\sigma \sim \mathcal{N}(0,1)$. Tabulated values of $\Phi(z)$ cover all Normal distributions.

Normal Distribution: Key Properties and the CLT

The 68–95–99.7 Rule (Empirical Rule):

- $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.6827$
- $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.9545$
- $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.9973$

Reproductive Property: If $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ are independent, then:

$$\sum_{i=1}^{n} a_i X_i \sim \mathcal{N}\left(\sum a_i \mu_i, \sum a_i^2 \sigma_i^2\right)$$

Central Limit Theorem: For i.i.d. $X_1, \ldots, X_n$ with mean μ and variance σ^2:

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as } n \to \infty$$

Critical values of $\mathcal{N}(0,1)$: $z_{0.90} = 1.282$, $z_{0.95} = 1.645$, $z_{0.975} = 1.960$, $z_{0.995} = 2.576$

Gamma Distribution

Setup: Waiting time until the α-th event in a Poisson process; generalizes the Exponential distribution.

Notation: $X \sim \mathrm{Gamma}(\alpha, \beta)$, $\quad \alpha > 0$ (shape), $\beta > 0$ (rate)

PDF:

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

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

CDF:

$$F(x) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}$$

(γ is the lower incomplete gamma function)

Mean:

$$\mu = \frac{\alpha}{\beta}$$

Variance:

$$\sigma^2 = \frac{\alpha}{\beta^2}$$

MGF:

$$M(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha}, \quad t < \beta$$

Special cases: $\alpha = 1$ gives $\mathrm{Exp}(\beta)$; $\alpha = k/2,\ \beta = 1/2$ gives $\chi^2(k)$.

Beta Distribution

Setup: Models random variables constrained to $[0, 1]$, such as probabilities, proportions, and rates.

Notation: $X \sim \mathrm{Beta}(\alpha,\ \beta), \quad \alpha, \beta > 0$

PDF:

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

where $B(\alpha, \beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

CDF:

$$F(x) = I_x(\alpha,\ \beta)$$

(regularized incomplete beta function)

Mean:

$$\mu = \frac{\alpha}{\alpha+\beta}$$

Variance:

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

MGF:

$$M(t) = 1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^k}{k!}$$

Notes: $\mathrm{Beta}(1, 1) = \mathrm{Uniform}(0, 1)$. Conjugate prior for Binomial/Bernoulli likelihood in Bayesian inference.

Chi-Squared Distribution

Setup: Sum of squares of k independent standard normal random variables; arises in inference about variances and goodness-of-fit tests.

Notation: $X \sim \chi^2(k), \quad k \in \{1, 2, 3, \ldots\}$ (degrees of freedom)

Definition: If $Z_1, \ldots, Z_k \sim \mathcal{N}(0, 1)$ i.i.d., then $X = Z_1^2 + \cdots + Z_k^2 \sim \chi^2(k)$.

PDF:

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

CDF:

$$F(x) = \frac{\gamma(k/2,\ x/2)}{\Gamma(k/2)}$$

Mean:

$$\mu = k$$

Variance:

$$\sigma^2 = 2k$$

MGF:

$$M(t) = (1 - 2t)^{-k/2}, \quad t < \tfrac{1}{2}$$

Applications: Goodness-of-fit (NIST SP 800-22), confidence intervals for σ^2, chi-square independence tests.

Python: Discrete Distributions

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

fig, axes = plt.subplots(1, 3, figsize=(12, 4))

# Binomial(20, 0.4)
k = np.arange(0, 21)
axes[0].bar(k, stats.binom.pmf(k, n=20, p=0.4), color='steelblue')
axes[0].set_title('Binomial(20, 0.4)'); axes[0].set_xlabel('k')

# Geometric(0.3)
k = np.arange(1, 20)
axes[1].bar(k, stats.geom.pmf(k, p=0.3), color='coral')
axes[1].set_title('Geometric(0.3)'); axes[1].set_xlabel('k')

# Poisson(4)
k = np.arange(0, 15)
axes[2].bar(k, stats.poisson.pmf(k, mu=4), color='seagreen')
axes[2].set_title('Poisson(4)'); axes[2].set_xlabel('k')
```

Python: Discrete Distributions (Cont.)

```python
plt.tight_layout(); plt.show()

# Verify mean and variance
print(f"Binomial:   mean={20*0.4:.1f}, var={20*0.4*0.6:.2f}")
print(f"Geometric:  mean={1/0.3:.3f}, var={0.7/0.09:.3f}")
print(f"Poisson:    mean=4, var=4")
```

```
Binomial:   mean=8.0, var=4.80
Geometric:  mean=3.333, var=7.778
Poisson:    mean=4, var=4
```

Python: Continuous Distributions

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

fig, axes = plt.subplots(1, 3, figsize=(12, 4))

# Exponential(lambda=1): PDF and CDF
x = np.linspace(0, 6, 400)
axes[0].plot(x, stats.expon.pdf(x, scale=1), 'b-', lw=2, label='PDF')
axes[0].plot(x, stats.expon.cdf(x, scale=1), 'r--', lw=2, label='CDF')
axes[0].set_title('Exp(1)'); axes[0].legend()

# Gamma(alpha=3, beta=1)
axes[1].plot(x, stats.gamma.pdf(x, a=3, scale=1), 'b-', lw=2)
axes[1].set_title('Gamma(3, 1)')

# Normal(mu=3, sigma=1): PDF and CDF
x2 = np.linspace(0, 6, 400)
axes[2].plot(x2, stats.norm.pdf(x2, loc=3, scale=1), 'b-', lw=2, label='PDF')
axes[2].plot(x2, stats.norm.cdf(x2, loc=3, scale=1), 'r--', lw=2, label='CDF')
axes[2].set_title('Normal(3, 1)'); axes[2].legend()
```

Python: Continuous Distributions

```
plt.tight_layout(); plt.show()

# Verify moments
print(f"Exp(1):       mean={stats.expon.mean():.2f}, var={stats.expon.var():.2f}")
print(f"Gamma(3,1):   mean={stats.gamma.mean(a=3):.2f}, var={stats.gamma.var(a=3):.2f}")
print(f"Normal(3,1): mean={stats.norm.mean(loc=3):.2f}, var={stats.norm.var(scale=1):.2f}
    ")
```

```
Exp(1):       mean=1.00, var=1.00
Gamma(3,1):   mean=3.00, var=3.00
Normal(3,1): mean=3.00, var=1.00
```

Python: MGF Verification via Moments

```python
import numpy as np
from scipy import stats
import sympy as sp

t = sp.Symbol('t')

# Poisson(lambda=3): MGF = exp(3*(exp(t)-1))
lam = 3
M_pois = sp.exp(lam * (sp.exp(t) - 1))
mean_sym  = sp.diff(M_pois, t).subs(t, 0)
E_X2_sym  = sp.diff(M_pois, t, 2).subs(t, 0)
var_sym   = E_X2_sym - mean_sym**2
print(f"Poisson(3) via MGF:  mean={mean_sym}, var={var_sym}")

# Normal(mu=2, sigma^2=5): MGF = exp(2t + 5t^2/2)
mu_n, sig2_n = 2, 5
M_norm = sp.exp(mu_n*t + sp.Rational(1,2)*sig2_n*t**2)
mean_n = sp.diff(M_norm, t).subs(t, 0)
var_n  = sp.diff(M_norm, t, 2).subs(t, 0) - mean_n**2
print(f"Normal(2,5) via MGF: mean={mean_n}, var={var_n}")
```

```
Poisson(3) via MGF:  mean=3, var=3
Normal(2,5) via MGF: mean=2, var=5
```

Discrete Distributions: Summary Table

Distribution	**Parameters**	**Mean**	**Variance**	**Support**
Bernoulli	p	p	$p(1-p)$	$\{0, 1\}$
Binomial	$n, \ p$	np	$np(1-p)$	$\{0, \dots, n\}$
Geometric	p	$1/p$	$(1-p)/p^2$	$\{1, 2, \dots\}$
Neg. Binomial	$r, \ p$	r/p	$r(1-p)/p^2$	$\{r, r+1, \dots\}$
Poisson	λ	λ	λ	$\{0, 1, 2, \dots\}$
Hypergeometric	N, K, n	nK/N	$n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$	$\{0, \dots, \min(n, K)\}$

Continuous Distributions: Summary Table

Distribution	**Parameters**	**Mean**	**Variance**	**Support**
Uniform	$a,\ b$	$(a+b)/2$	$(b-a)^2/12$	$[a, b]$
Exponential	λ	$1/\lambda$	$1/\lambda^2$	$[0, \infty)$
Normal	$\mu,\ \sigma^2$	μ	σ^2	$(-\infty, \infty)$
Gamma	$\alpha,\ \beta$	α/β	α/β^2	$(0, \infty)$
Beta	$\alpha,\ \beta$	$\alpha/(\alpha+\beta)$	see slide	$[0, 1]$
Chi-Squared	k	k	$2k$	$[0, \infty)$

MGF Reference Table

Distribution	**MGF** $M(t)$	**Domain**
Bernoulli(p)	$1 - p + pe^t$	all t
Binomial(n, p)	$(1 - p + pe^t)^n$	all t
Geometric(p)	$pe^t \,/\, (1 - (1 - p)e^t)$	$t < -\ln(1 - p)$
Neg. Binomial(r, p)	$(p \,/\, (1 - (1 - p)e^t))^r$	$t < -\ln(1 - p)$
Poisson(λ)	$\exp(\lambda(e^t - 1))$	all t
Uniform(a, b)	$(e^{tb} - e^{ta}) \,/\, (t(b - a))$	all $t \neq 0$
Exponential(λ)	$\lambda \,/\, (\lambda - t)$	$t < \lambda$
Normal(μ, σ^2)	$\exp(\mu t + \sigma^2 t^2/2)$	all t
Gamma(α, β)	$(1 - t/\beta)^{-\alpha}$	$t < \beta$
Chi-Squared(k)	$(1 - 2t)^{-k/2}$	$t < 1/2$

Relationships Between Distributions

- **Bernoulli $\rightarrow$ Binomial:** Sum of n independent Bernoulli(p) $\equiv$ Binomial(n, p)
- **Binomial $\rightarrow$ Poisson:** $n \rightarrow \infty,\ p \rightarrow 0,\ np \rightarrow \lambda$: Binomial($n, p$) $\rightarrow$ Poisson(λ)
- **Geometric $\rightarrow$ Neg. Binomial:** Sum of r independent Geometric(p) $\equiv$ NegBin(r, p)
- **Exponential $\rightarrow$ Gamma:** Sum of α independent Exp(β) $\equiv$ Gamma(α, β)
- **Normal $\rightarrow$ Chi-Squared:** Sum of squares of k i.i.d. $\mathcal{N}(0,1) \equiv \chi^2(k)$
- **Gamma $\rightarrow$ Exponential:** Gamma($1, \beta$) $\equiv$ Exp(β)
- **Gamma $\rightarrow$ Chi-Squared:** Gamma($k/2,\ 1/2$) $\equiv \chi^2(k)$
- **Beta $\rightarrow$ Uniform:** Beta($1, 1$) $\equiv$ Uniform($0, 1$)
- **Binomial $\rightarrow$ Normal (CLT):** Binomial(n, p) $\approx \mathcal{N}(np,\ np(1-p))$ for large n
- **Hypergeometric $\rightarrow$ Binomial:** As $N \rightarrow \infty$ with $K/N \rightarrow p$: Hypergeometric $\rightarrow$ Binomial(n, p)

Summary and Key Takeaways

Core Concepts

- PMF for discrete; PDF for continuous
- CDF unifies both: $F(x) = P(X \leq x)$
- Mean $=$ probability-weighted center of mass
- Variance $=$ average squared deviation from mean
- MGF encodes all moments and *uniquely* identifies the distribution

Always check:

- Parameterization (rate vs. scale)
- Support set
- Existence of MGF (heavy-tailed dists may lack it)

Distributions in Practice

- **Poisson / Exponential:** Arrival and wait-time modeling
- **Binomial:** Quality control, hypothesis tests
- **Normal:** CLT limit, measurement error, test statistics
- **Gamma:** Reliability, queuing, Bayesian models
- **Beta:** Proportions, Bayesian priors (conjugate for Binomial)
- **Chi-Squared:** Goodness-of-fit, variance inference

Golden Rule

Verify which parameterization a software library uses before computing probabilities numerically (e.g.,

Questions?