Table 1: Common Discrete Distributions

Distribution	Parameters	PMF ($\mathbb{P}(X = k)$)	CMF ($\mathbb{P}(X \leq k)$)	Expectation ($\mathbb{E}[X]$)	Variance ($\mathrm{Var}(X)$)	Support
Uniform	$\mathrm{Uniform}(a, b)$	$\frac{1}{b-a+1}$	$\frac{k-a+1}{b-a+1}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$	$X \in [a, b]$
Bernoulli	$\mathrm{Bernoulli}(p)$	$\begin{cases} 1 & p \\ 0 & 1-p \end{cases}$	—	p	$p(1-p)$	$X \in \{0, 1\}$
Binomial	$\mathrm{Bin}(n, p)$	$\binom{n}{k} p^k (1-p)^{n-k}$	—	np	$np(1-p)$	$X \in \mathbb{N}$
Geometric	$\mathrm{Geom}(p)$	$p(1-p)^{k-1}$	$1-(1-p)^k$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$X \in \mathbb{N}$
Poisson	$\mathrm{Pois}(\lambda)$	$\frac{\lambda^k e^{-\lambda}}{k!}$	—	λ	λ	$X \in \mathbb{N}$
Hypergeometric	$\mathrm{Hypergeometric}(N, K, n)$	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	—	$n\frac{K}{N}$	$n\frac{K(N-K)(N-n)}{N^2(N-1)}$	$X \in \mathbb{N}$

Table 2: Common Continuous Distributions

Distribution	Parameters	PDF ($f_X(x)$)	CDF ($F_X(x) = \mathbb{P}(X \leq x)$)	Expectation ($\mathbb{E}[X]$)	Variance ($\mathrm{Var}(X)$)	Support
Uniform	$\mathrm{Uniform}(a, b)$	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$X \in [a, b]$
Exponential	$\mathrm{Exp}(\lambda)$	$\lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$X \in [0, \infty)$
Normal/Gaussian	$\mathcal{N}(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\Phi(x)$	μ	σ^2	$X \in \mathbb{R}$

1 Discrete Distribution Properties

1.1 Binomial Distribution

- The binomial distribution represents the probability of getting k successes from n draws, where each draw independently has a probability p of being a success.
- It's usually helpful to think of binomial distributions as a sum of n iid Bernoulli(p) RVs. This is how the expectation and variance formulas are derived.

1.2 Geometric Distribution

- Geometric distributions model the number of trials we need before we get a success (including the success), where each trial is a success with probability p.
- There are two kinds of geometric distributions; one where we count failures, and one where we count trials (i.e. failures *and* the one success).

  The formula listed in the table counts trials; $X = k$ means that we have $k - 1$ failures and 1 success. If we wanted to count failures instead, we'd just substitute $k + 1$ instead of k:

  $$\mathbb{P}(X = k) = p(1-p)^k \qquad \mathbb{P}(X \leq k) = 1 - (1-p)^{k+1} \qquad \mathbb{E}[X] = \frac{1-p}{p} = \frac{1}{p} - 1$$

  We can express this distribution of counting failures as $X \sim \text{Geom}(p) - 1$. This corresponds directly to counting trials, and subtracting the one success.

  It's important to know what exactly you're counting when dealing with problems involving geometric distributions, as the formulas and results are different in the end.

- **Memoryless Property**:

  If $X \sim \text{Geom}(p)$, then the memoryless property gives

  $$\mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s).$$

  In other words, the probability that we'd have to wait some time s before a success occurs is the same no matter when we start observing.

1.3 Poisson Distribution

- Poisson distributions model the number of events that occur in a fixed time interval; the parameter λ represents the average number of occurrences in that fixed time interval.
- **Sum of Independent Poissons**:

  If $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$ are independent, then $X + Y \sim \text{Pois}(\lambda + \mu)$.

  More generally, if $X_i \sim \text{Pois}(\lambda_i)$, then $\sum_{i=1}^n X_i \sim \text{Pois}\left(\sum_{i=1}^n \lambda_i\right)$.

- **Poisson as the Limit of Binomial**:

  If $X \sim \text{Bin}(n, \frac{\lambda}{n})$, then $\mathbb{P}(X = k) \to \frac{\lambda^k e^{-\lambda}}{k!}$ as $n \to \infty$.

  In other words, as $n \to \infty$, $\text{Bin}(n, \frac{\lambda}{n}) \to \text{Pois}(\lambda)$.

1.4 Hypergeometric Distribution

- Similar to the binomial distribution, the hypergeometric distribution describes the probability of k successes in n draws, but *without* replacement and from a population of N items, of which K items count as successes.

  Hypergeometric distributions are very uncommon in CS70 problems, so most of the time, you'd really only be asked to identify the distribution, and perhaps give the expectation (derived from linearity of expectation with indicators).

2 Continuous Distribution Properties

When working with continuous distributions, it's usually best to work in terms of the CDFs. This is because densities are often cumbersome to work with, and it's usually a lot easier to get the PDF from a CDF (by differentiating), rather than get a CDF from a PDF (by integrating), as differentiation is a lot simpler than integration.

2.1 Exponential distribution

- Exponential distributions are the continuous analog to the geometric distribution; it models the time we have to wait until we observe an event, when the event occurs with rate λ.
- **Memoryless Property**:

  If $X \sim \text{Exp}(\lambda)$, then the memoryless property gives

  $$\mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s).$$

  This is exactly the same as the memoryless property for geometric distributions.
- **Minimum of Exponentials**:

  If $X \sim \text{Exp}(\lambda_1)$ and $Y \sim \text{Exp}(\lambda_2)$, then

  $$\min(X, Y) \sim \text{Exp}(\lambda_1 + \lambda_2).$$

  This can be extended to multiple random variables as well; the minimum is exponential with the sum of the rates. This is derived from finding $\mathbb{P}(\min X_i > x)$ and simplifying.

2.2 Normal distribution

- To standardize a normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$, we have

  $$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

- **Central Limit Theorem**:

  If we have n iid samples X_i with mean μ and variance σ^2, let $\frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample average. We then have $\mathbb{E}[\frac{1}{n}S_n] = \mu$ and $\text{Var}(\frac{1}{n}S_n) = \frac{1}{n}\sigma^2$.

  The central limit theorem states that

  $$\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{n\to\infty} \mathcal{N}(0, 1).$$

  Using this, we can create confidence intervals or find probabilities for S_n or $\frac{1}{n}S_n$ by approximating the sample average with a normal distribution and using the normal CDF.