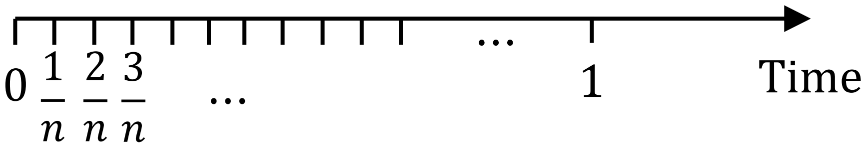
CS70 @ UC Berkeley, Spring 2026

Lecture 24
Continuous Probability Distribution I
April 21, 2026

Poisson Approximation of Binomial(n, p)

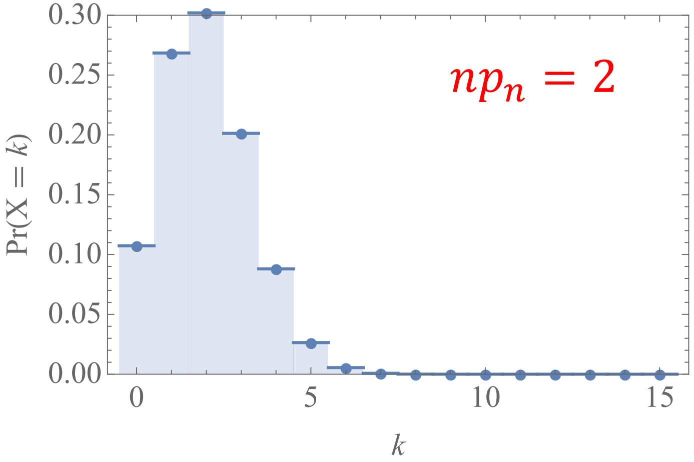
- Consider i.i.d. Bernoulli(p_n) trials, with 1 trial for each $\frac{1}{n}$ interval.



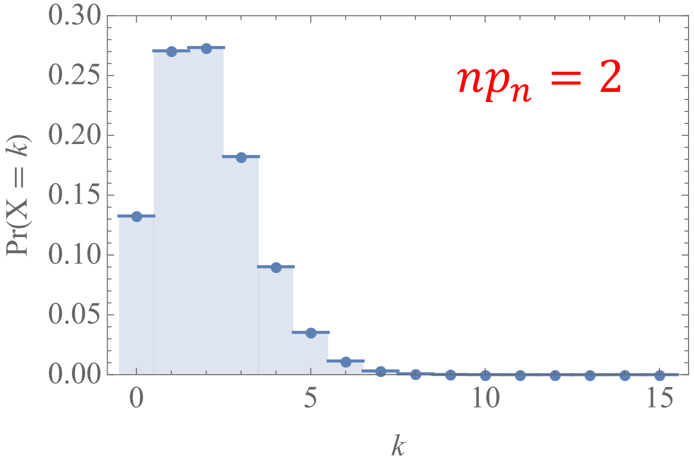
- Consider a sequence $X_k, X_{k+1}, X_{k+2}, \ldots$ of RVs where $X_n \sim$ Binomial(n, p_n).
- (∗) Limit as $n \to \infty$ and $p_n \to 0$ while $np_n = \lambda$ is held fixed.
- In Lecture 22, we saw that the distribution of X_n converges to Poisson(λ) as $n \to \infty$.
- **Aside:** in more advanced probability theory, this kind of convergence of random variables is written as $X_n \xrightarrow{d} X \sim$ Poisson(λ) as $n \to \infty$ and we say that the sequence $\{X_n\}$ converges in distribution to X as n tends to ∞.

This corresponds to a continuous time limit

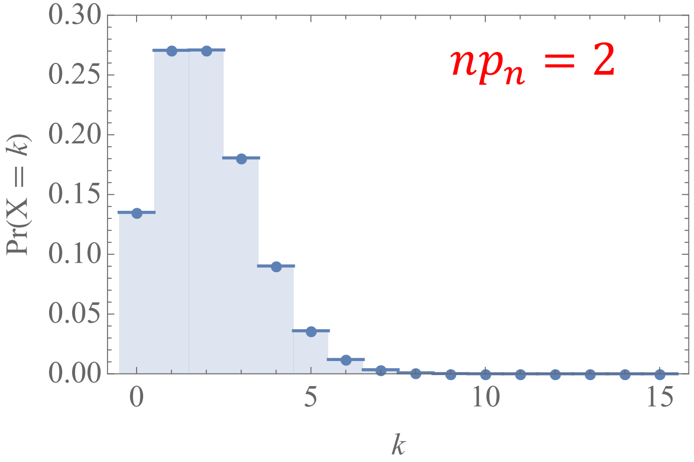
Binomial with $n = 10$, $p = 0.2$



Binomial with $n = 100$, $p = 0.02$

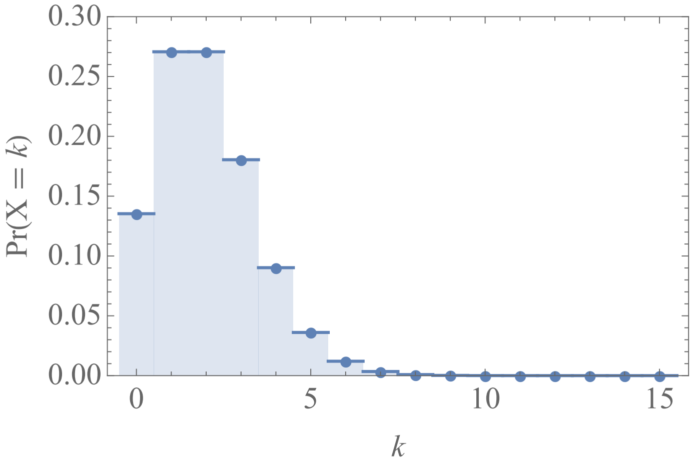


Binomial with $n = 1000$, $p = 0.002$



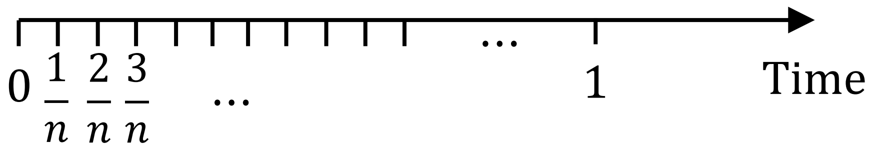
These two distributions are virtually indistinguishable

$\lambda = 2$



Continuous Time Limit of Geometric(p)

- Consider i.i.d. Bernoulli(p_n) trials, with 1 trial for each $\frac{1}{n}$ interval.



- T_n = physical waiting time to the first success. $T_n \sim \text{Geometric}(p_n)$
- $\mathbb{P}(T_n > k/n) = (1 - p_n)^k$
- For any $t \in \mathbb{R}_+$, $\mathbb{P}(T_n > t) = \mathbb{P}\left(T_n > \lfloor tn \rfloor \frac{1}{n}\right) = (1 - p_n)^{\lfloor tn \rfloor}$
- $(*)$ Limit as $n \to \infty$ and $p_n \to 0$ while $np_n = \lambda > 0$ is held fixed.
- $\mathbb{P}(T_n > t) = \left(1 - \frac{\lambda}{n}\right)^{\lfloor tn \rfloor} \to e^{-\lambda t}$ in the limit $(*)$.
- This limiting distribution is called the Exponential Distribution
- We write $T_n \xrightarrow{d} T \sim \text{Exp}(\lambda)$ as $n \to \infty$.
- $\mathbb{P}(T \le t) = 1 - \mathbb{P}(T > t) = 1 - e^{-\lambda t}$ (This is called the Cumulative Distribution Function.)

Cumulative Distribution Function (CDF)

Definition (CDF). Given a random variable X, its cumulative distribution function F_X is defined as

$$F_X(a) = \mathbb{P}(X \le a), \text{for } a \in (-\infty, +\infty)$$

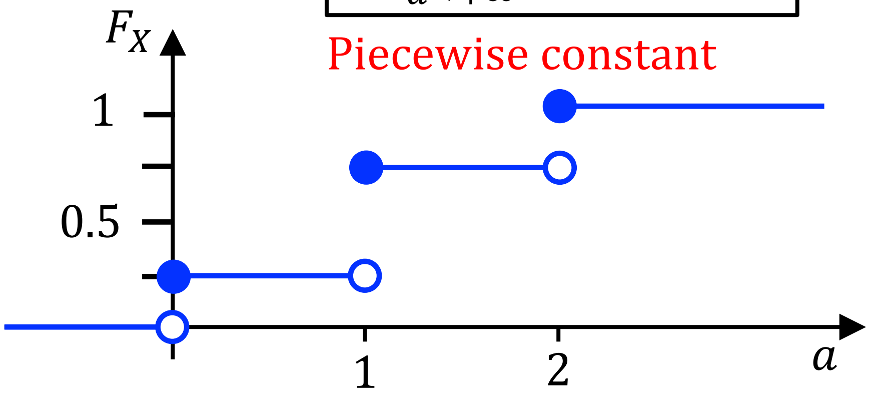
Discrete Random Variable X:

- $F_X(a) = \sum_{b:b\le a} \mathbb{P}(X = b)$

- Non-decreasing
- Right-continuous
- $\lim_{a\to-\infty} F_X(a) = 0$
- $\lim_{a\to+\infty} F_X(a) = 1$

Example: Toss a fair coin twice.
$X(\omega) =$ Heads in $\omega \in \Omega$.

$X = 0: \{(T,T)\}$
$X = 1: \{(T,H), (H,T)\}$
$X \le 1: \{(T,T), (T,H), (H,T)\}$
$X = 2: \{(H,H)\}$
$X \le 2: \{(T,T), (T,H), (H,T), (H,H)\}$



Piecewise constant

Continuous Distribution

Continuous Random Variable X:

- $F_X(a) = \mathbb{P}(X \leq a)$ is continuous $\forall a \in \mathbb{R}$.
- We will consider continuous random variables with densities such that:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$$

Probability Density Function (PDF)

- $\mathbb{P}(\underbrace{-\infty \leq X \leq \infty}_{\Omega}) = \int_{-\infty}^{\infty} f_X(x)dx = 1$
- In particular, $F_X(a) = \int_{-\infty}^{a} f_X(x)dx$
- If F is differentiable at x, $\frac{dF_X(x)}{dx} = f_X(x)$

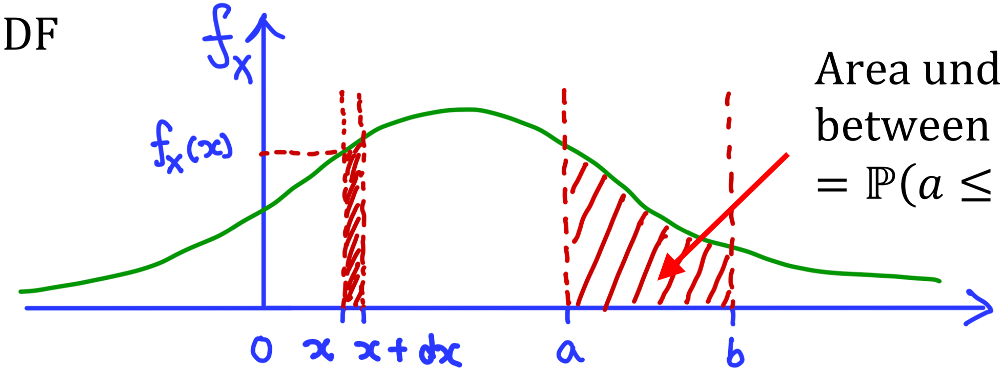
Example: Exponential Distribution

CDF $F_T(t) = \mathbb{P}(T \leq t) = 1 - \mathbb{P}(T > t) = 1 - e^{-\lambda t}$

So, the PDF of T is $f_T(t) = \frac{dF_T(t)}{dt} = \lambda e^{-\lambda t}$

Continuous Distribution

PDF



Area under the curve between a and b $= \mathbb{P}(a \le X \le b)$

- Suppose X is a continuous random variable. Then, $\mathbb{P}(X = a) = 0$ for all $a \in \mathbb{R}$
- $\mathbb{P}(X = a) = \lim_{\delta \to 0} \mathbb{P}\,(a \le X \le a + \delta) = \lim_{\delta \to 0} \int_a^{a+\delta} f_X(x)dx$

  $= \lim_{\delta \to 0} [F_X(a+\delta) - F(a)] = 0$

  By the Fundamental Theorem of Calculus

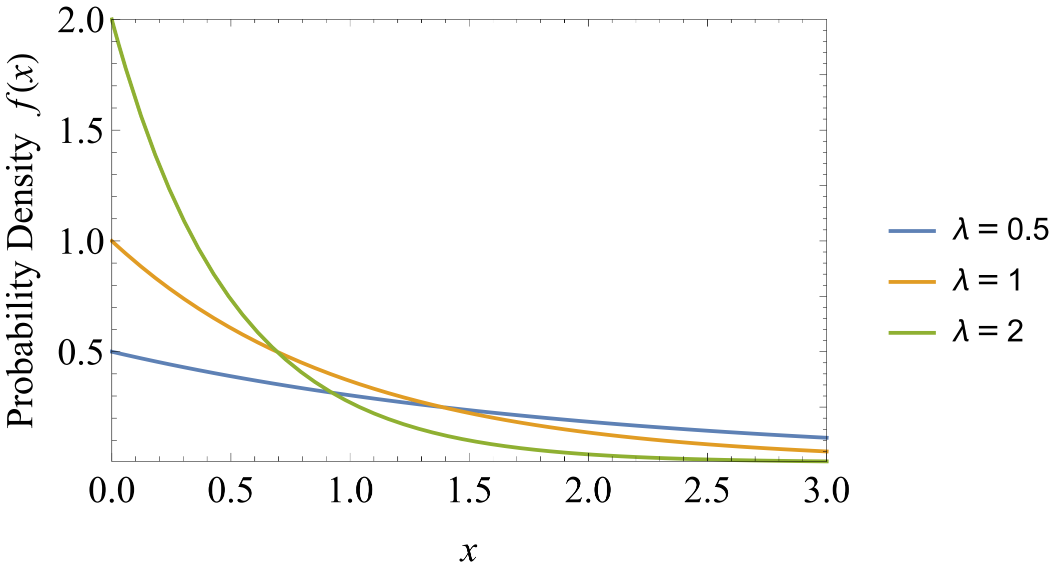
  since F_X is a continuous function
- Infinitesimal probability (from Riemann Integral): For $dx \ll 1$

$$\mathbb{P}(x \le X \le x + dx) \approx f_X(x)dx$$

- **Expectation** $\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)dx$
- **Variance** $\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x)dx$

Exponential Distribution

- $T \sim \text{Exp}(\lambda)$, where $\lambda > 0$.
- PDF $f_T(t) = \frac{dF_T(t)}{dt} = \lambda e^{-\lambda t}$

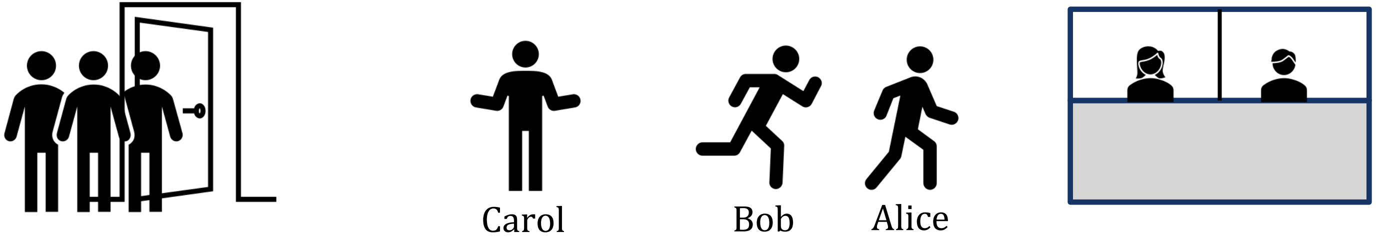


- $\mathbb{E}[T] = \int_{-\infty}^{\infty} t\, f_T(t) dt = \frac{1}{\lambda}$
- $\text{Var}[T] = \int_{-\infty}^{\infty} (t - \mathbb{E}[T])^2 f_T(t) dt = \frac{1}{\lambda^2}$.
- **Memoryless property:** For all $t, s > 0$,

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

Who's Last?

- **Alice**, **Bob** and **Carol** arrive at the post office at the same time. There are only two counters, and **Alice** and **Bob** rush to take them.
- Assume that the service time per customer is distributed as $\mathrm{Exp}(\lambda)$, for some $\lambda > 0$.
- What is the probability that **Carol** is the last one of the three to be done with service?



Answer: 1/2

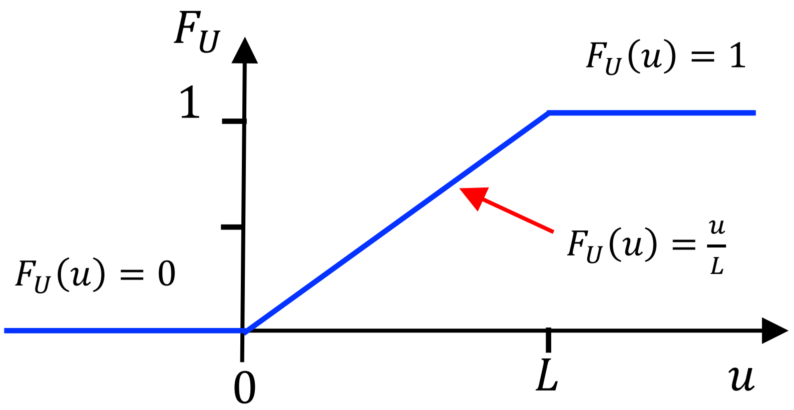
- Case 1: Alice is done before Bob with probability ½. Given this event, Carol joins Bob at the counter, and by the memoryless property of the exponential distribution, their service times are independent $\mathrm{Exp}(\lambda)$ RVs, and, by symmetry, Carol finishes after Bob with probability ½.
- Case 2: Bob is done before Alice with probability ½. Given this event, a similar argument as above shows that Carol finishes after Alice with probability ½.
- Combining everything using the Law of Total Probability, we obtain answer = ½.

Continuous Uniform Distribution

Uniform Distribution $U \sim \text{Uniform}[0, L]$

$$\mathbb{P}(a \leq U \leq b) = \frac{b-a}{L} \text{ for } a, b \in [0, L], a < b$$

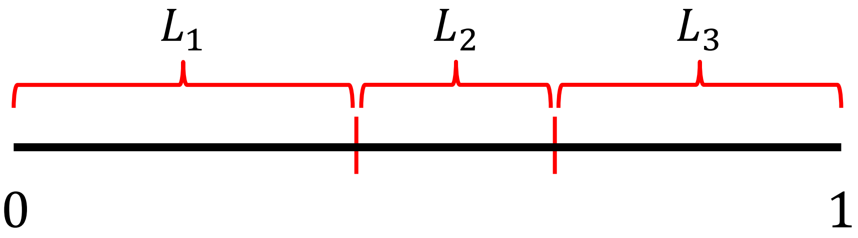
$$f_U(u) = \begin{cases} \frac{1}{L}, & \text{for } 0 \leq u \leq L \\ 0, & \text{otherwise} \end{cases}$$



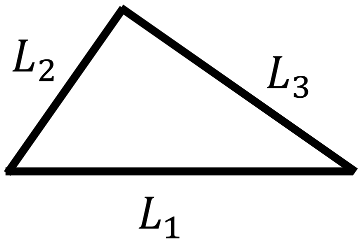
- $\mathbb{E}[U] = \int_{-\infty}^{\infty} u\, f_U(u) du = \int_0^L u \frac{1}{L} du = \frac{L}{2}$
- $\text{Var}[U] = \int_{-\infty}^{\infty} (u - \mathbb{E}[U])^2 f_U(u) du = \frac{L^2}{12}$.

Random Breaks to Form a Triangle

- Suppose Alice samples two points independently and uniformly at random from the interval [0,1], thereby obtaining 3 segments.



- What is the Probability that the three segments can form a triangle?



The answer will be provided in the next lecture.

Joint Density

Definition (Joint Density): A joint density function for two random variables on the same probability space is a function $f_{X,Y}: \mathbb{R}^2 \to \mathbb{R}$:

1. $f_{X,Y}(x, y) \geq 0, \forall x, y \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \mathrm{d}x\mathrm{d}y = 1$.

For all well-defined $A \subset \mathbb{R}^2$,

$$\mathbb{P}[(X, Y) \in A] = \int \int_A f_{X,Y}(x, y) \mathrm{d}x\mathrm{d}y$$

Consider a small neighborhood Δ containing (x, y). Then,

$$\mathbb{P}[(X, Y) \in \Delta] \approx f_{X,Y}(x, y) \mathrm{Area}(\Delta)$$

Marginal and Conditional Density

Definition (Marginal Density). Consider a joint distribution on X and Y with joint density function $f_{X,Y}(x,y)$.
The marginal density function f_X of X is defined as $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y$.
The marginal density function f_Y of Y is defined as $f_Y(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x$.

Given $A \subset \mathbb{R}$, how should we define $\mathbb{P}(X \in A \mid Y = b)$?
Subtlety: $\mathbb{P}(Y = b) = 0, \forall b \in \mathbb{R}$. For $\delta \ll 1$,

$$\mathbb{P}(X \in A \mid b < Y < b+\delta) = \frac{\mathbb{P}(X \in A,\ b < Y < b+\delta)}{\mathbb{P}(b < Y < b+\delta)}$$

$$= \frac{\int_A \int_b^{b+\delta} f_{X,Y}(x,y)\mathrm{d}y\,\mathrm{d}x}{\int_b^{b+\delta} f_Y(y)\mathrm{d}y} \approx \frac{\int_A f_{X,Y}(x,b)\delta\,\mathrm{d}x}{f_Y(b)\delta}$$

$$= \int_A \boxed{\frac{f_{X,Y}(x,b)}{f_Y(b)}}\mathrm{d}x$$

Conditional density of X given $Y = b$, denoted $f_{X|Y=b}(x)$, which is well defined as long as $f_Y(b) > 0$.

Independence and the Law of Total Probability

Independence:

X, Y independent $\Leftrightarrow f_{X|Y=y}(x) = f_X(x), \forall x, y \in \mathbb{R}$ such that $f_Y(y) > 0$

$$\Leftrightarrow f_{X,Y}(x,y) = f_X(x) f_Y(y), \forall x, y \in \mathbb{R}$$

Theorem (Law of Total Probability or Total Probability Rule), Lecture 17.
Suppose $B_1, B_2, B_3, \dots,$ is a partition of Ω. Then, for any $A \subseteq \Omega$,

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

a.k.a. rule of average conditional probabilities.

Law of Total Probability (Continuous Case):

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{dy} = \int_{-\infty}^{\infty} f_{X|Y=y}(x) f_Y(y)\mathrm{dy}$$