ENGG 2430 / ESTR 2004: Probability and Statistics

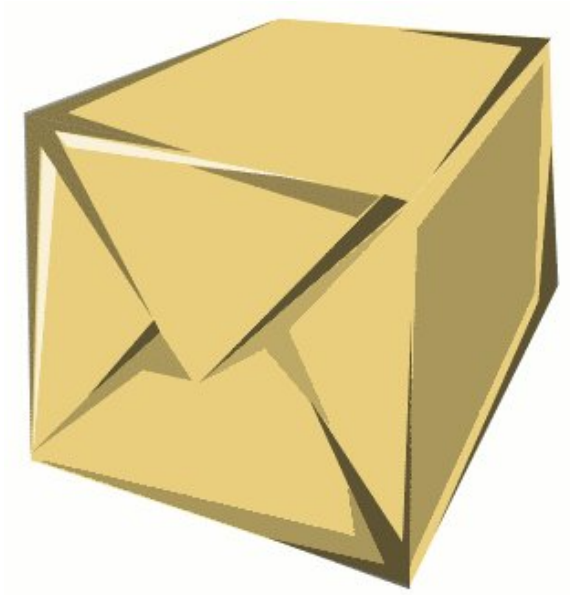
Spring 2019

6. Continuous Random Variables I

Andrej Bogdanov

Delivery time

A package is to be delivered between noon and 1pm.



What is the expected arrival time?

12.30

Discrete model I

$\Omega = \{0, 1, \ldots, 59\}$

equally likely outcomes

X**: minute when package arrives**

$$E[X] = 0 \cdot \frac{1}{60} + 1 \cdot \frac{1}{60} + \ldots + 59 \cdot \frac{1}{60} = \underline{29.5}$$

Discrete model II

$$\Omega = \{0, {}^{1}\!/\!{}_{60}, {}^{2}\!/\!{}_{60}, \ldots, 1, 1\,{}^{1}\!/\!{}_{60}, \ldots, 59\,{}^{59}\!/\!{}_{60}\}$$

equally likely outcomes

X: **minute when package arrives**

$$E[X] = 0 \cdot \frac{1}{60^2} + \frac{1}{60} \cdot \frac{1}{60^2} + \ldots + \left(59\frac{59}{60}\right)\frac{1}{60^2}$$

$$= 29.9\ldots$$

Continuous model

$\Omega =$ **the (continuous) interval** $[0, 60)$

equally likely outcomes

X: **minute when package arrives**

$P(X = 35.62) = 0$

$P(X = 30) = 0$

Uncountable sample spaces

In Lecture 2 we said:

> "*The probability of an event is the sum of the probabilities of its elements*"

but in $[0, 60)$ all elements have probability zero!

To specify and calculate probabilities, we have to work with the axioms of probability

The uniform random variable

Sample space $\Omega = [0, 60)$

Events of interest: **intervals** $[x, y) \subseteq [0, 60)$

their intersections, unions, etc.

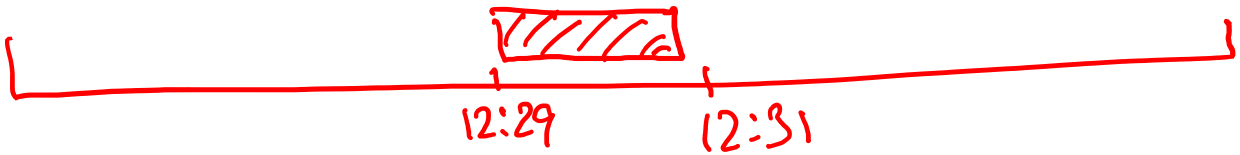
Probabilities: $P([x, y)) = (y - x)/60$

Random variable: $X(\omega) = \omega$

$P(X \leq 31) = \frac{31}{60}$

$P(X \leq 29) = \frac{29}{60}$

$P(29 \leq X < 31) = \frac{2}{60} = \frac{1}{30}.$



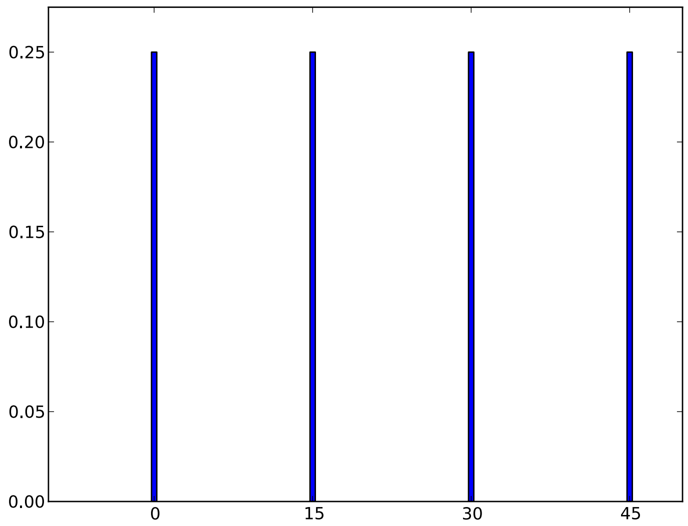
Cumulative distribution function

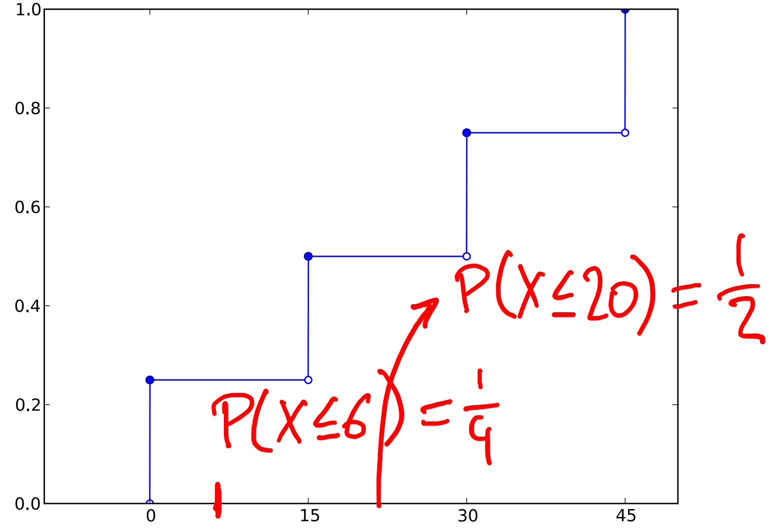
The probability mass function doesn't make much sense because $P(X = x) = 0$ for all x.

Instead, we can describe X by its cumulative distribution function (CDF) F:

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

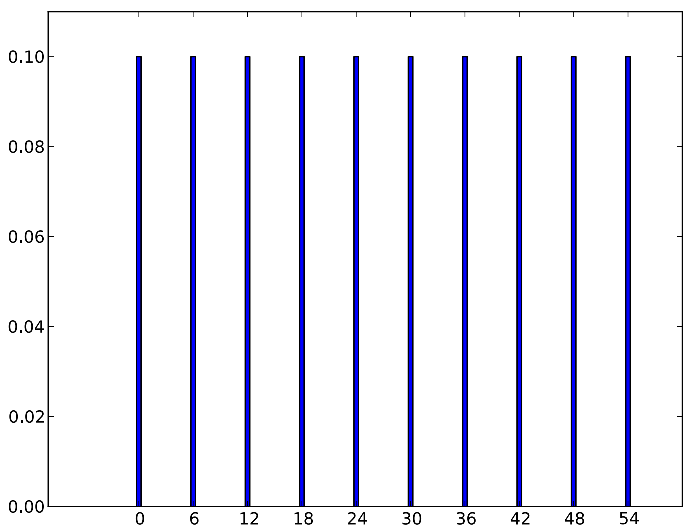
Cumulative distribution functions

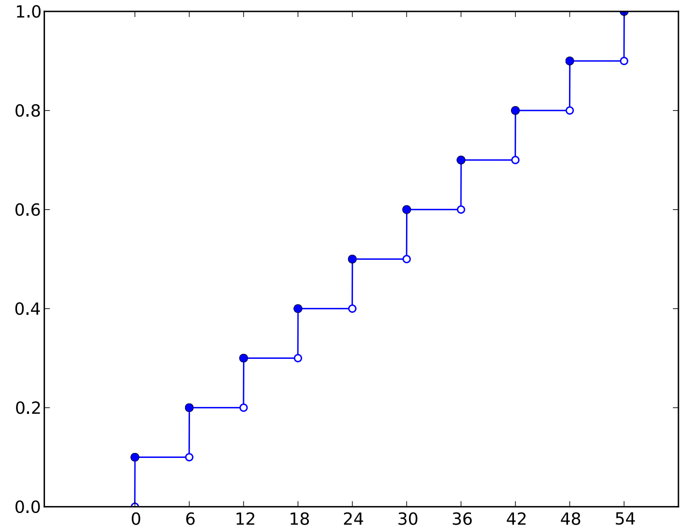




$$f_X(x) = P(X = x)$$

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



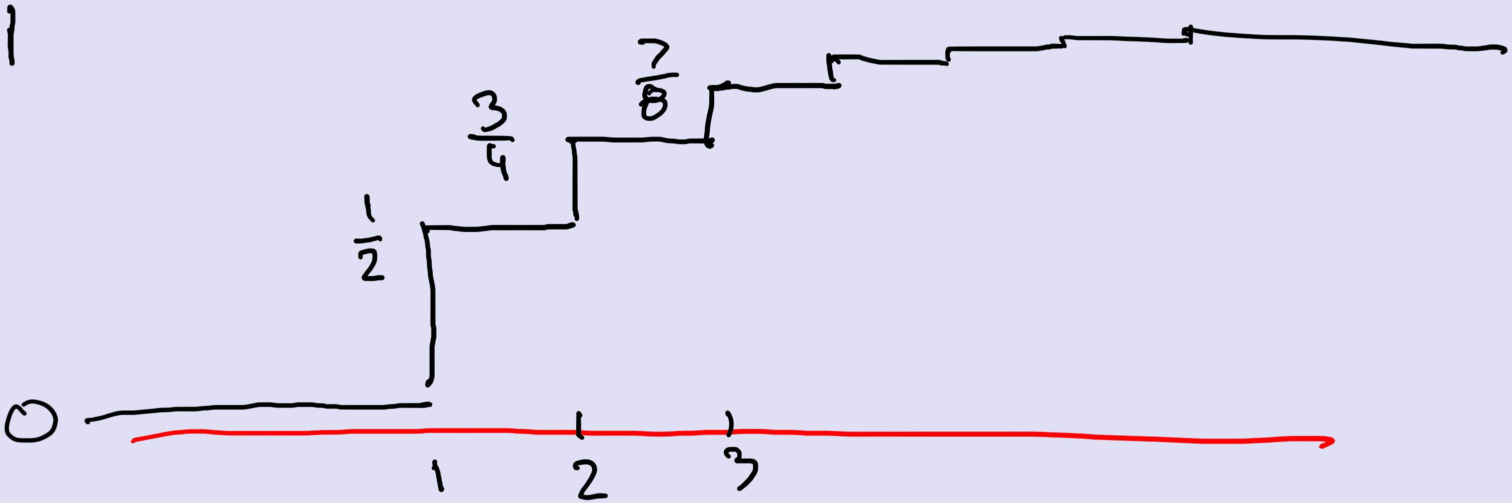


What is the Geometric$(1/2)$ CDF?

PDF

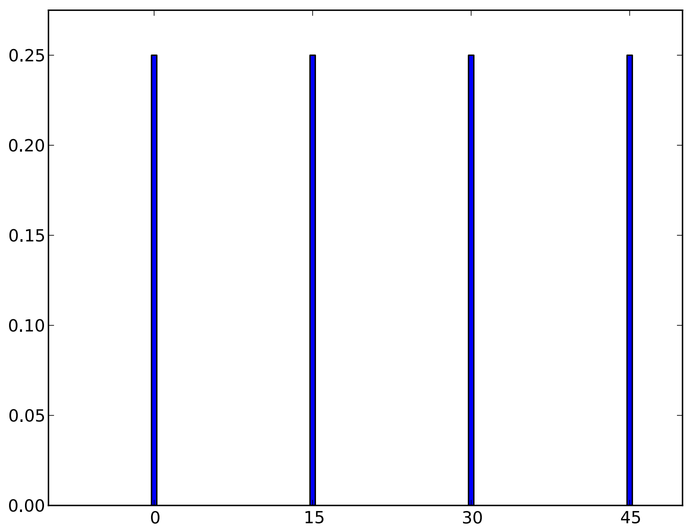
x	1	2	3	. . .
$P(X=x)$	$1/2$	$1/4$	$1/8$	. . .

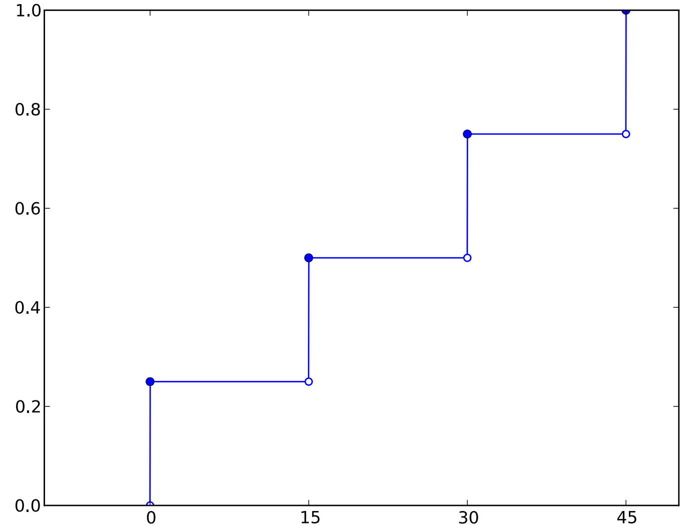
CDF



$$P(X \le k) = 1 - P(X > k)$$
$$= 1 - P(\text{FIRST } k \text{ TRIALS FAILED})$$
$$= 1 - 1/2^k.$$

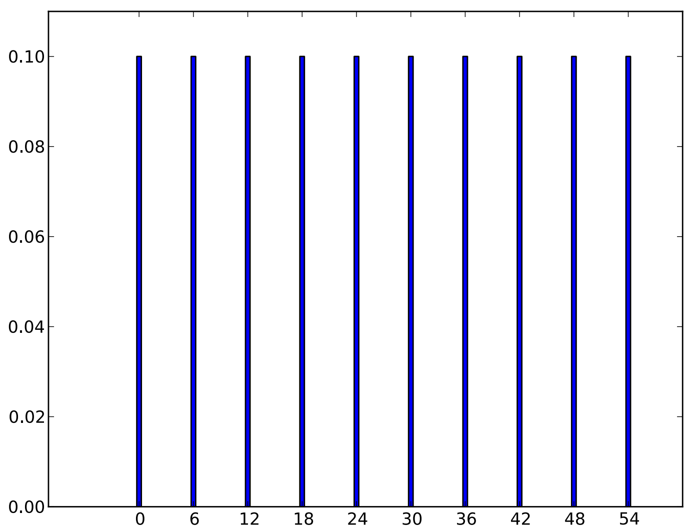
Cumulative distribution functions

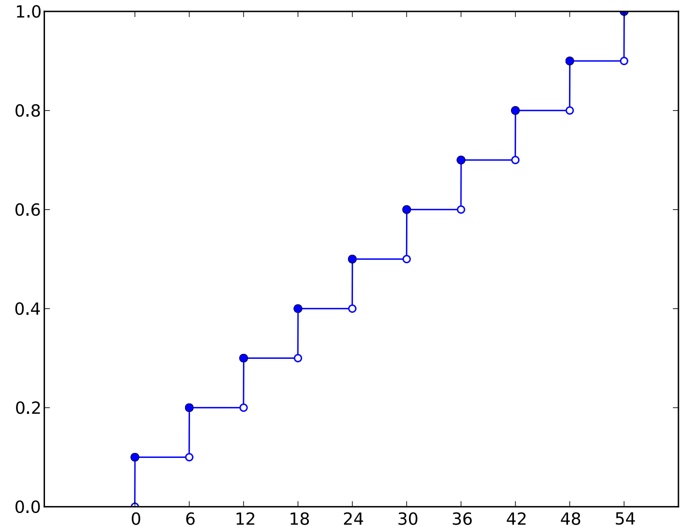




$$f(x) = P(X = x)$$

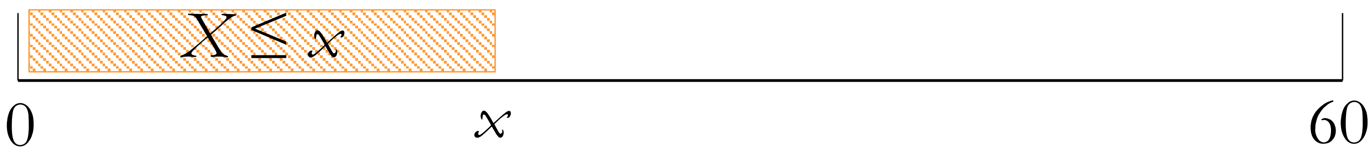
$$F(x) = P(X \leq x)$$



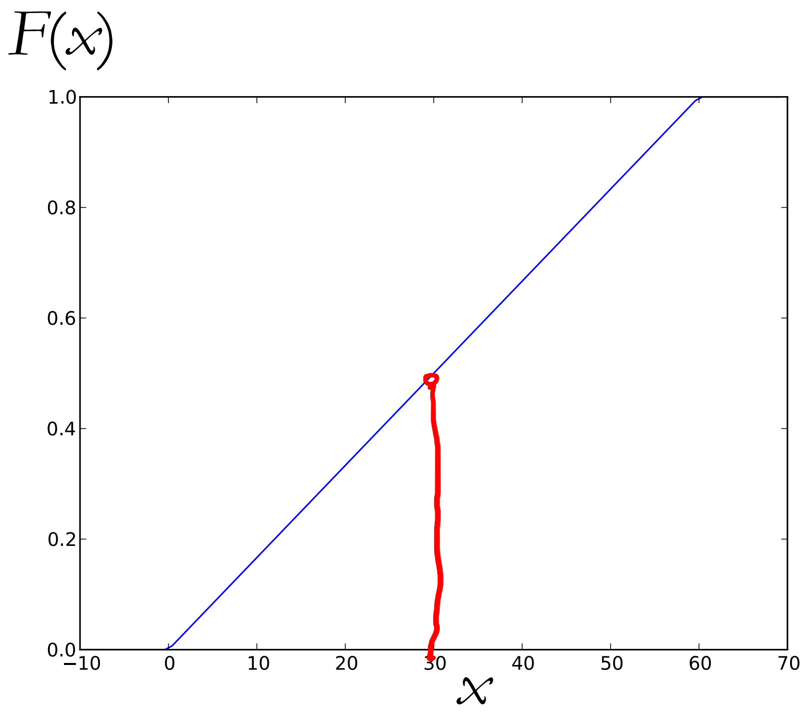


Uniform random variable

If X is uniform over $[0, 60)$ then

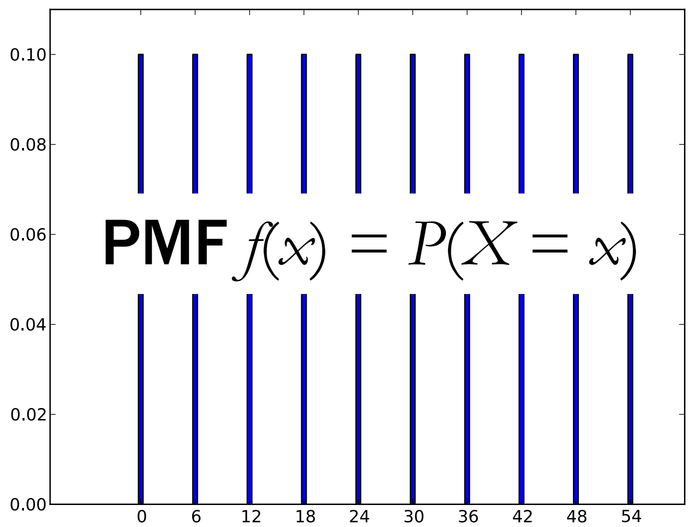


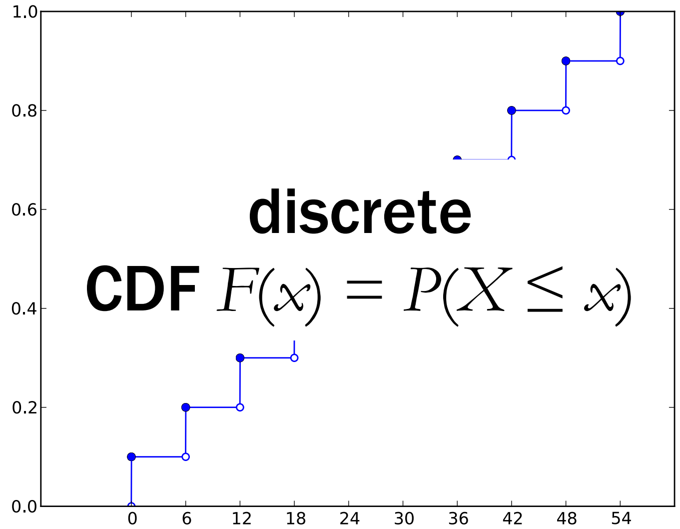
$$P(X \le x) = \begin{cases} 0 & \textbf{for } x < 0 \\ x/60 & \textbf{for } x \in [0, 60) \\ 1 & \textbf{for } x > 60 \end{cases}$$



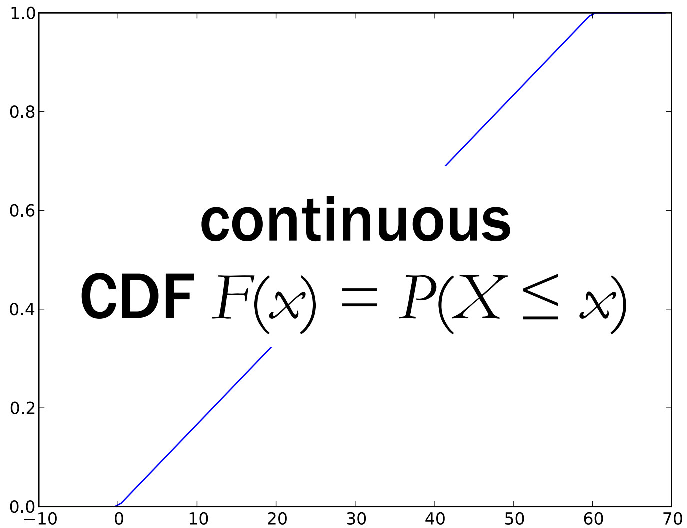
$P(X \le 30) = \frac{30}{60} = \frac{1}{2}$

Cumulative distribution functions





?



Discrete random variables:

PMF $f(x) = P(X = x)$

CDF $F(x) = P(X \leq x)$

$f(x) = F(x) - F(x - \delta)$

for small δ

$F(a) = \sum_{x \leq a} f(x)$

Continuous random variables:

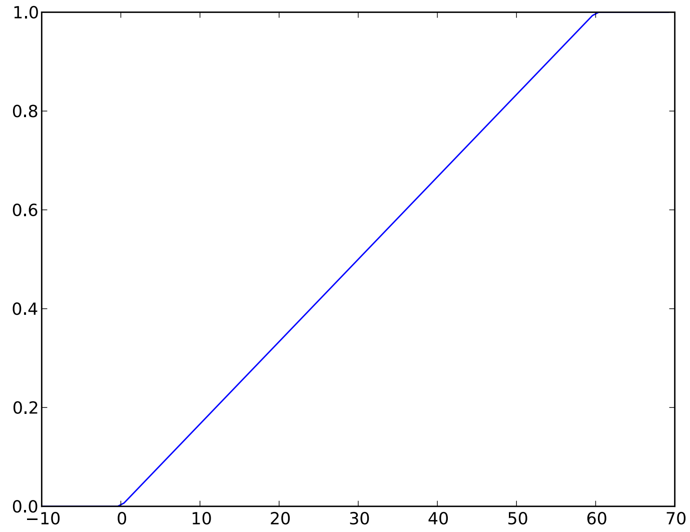
The probability density function (PDF) of a random variable with CDF $F(x)$ **is**

$$f(x) = \lim_{\delta \to 0} \frac{F(x) - F(x - \delta)}{\delta} = \frac{dF(x)}{dx}$$

Uniform random variable

CDF

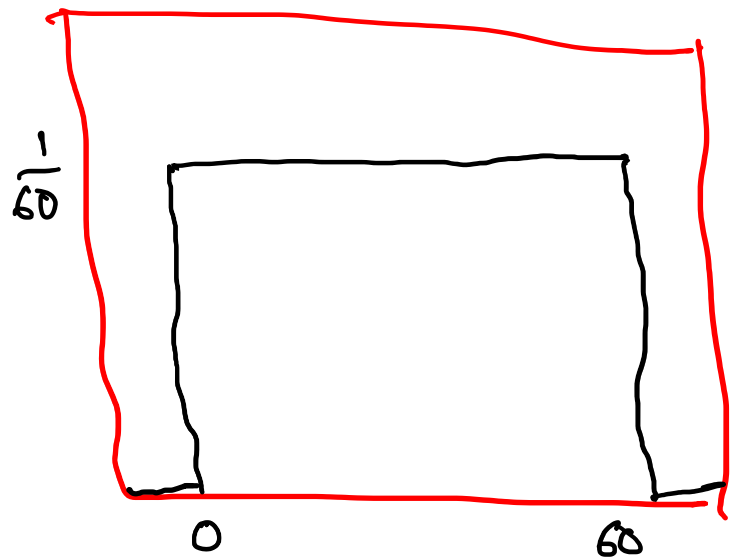
$$F(x) = \begin{cases} 0 & \textbf{if } x < 0 \\ x/60 & \textbf{if } x \in [0, 60) \\ 1 & \textbf{if } x \geq 60 \end{cases}$$



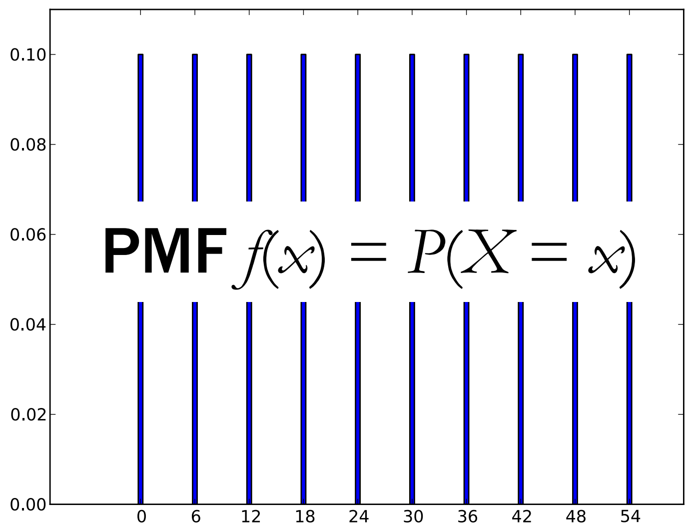
PDF

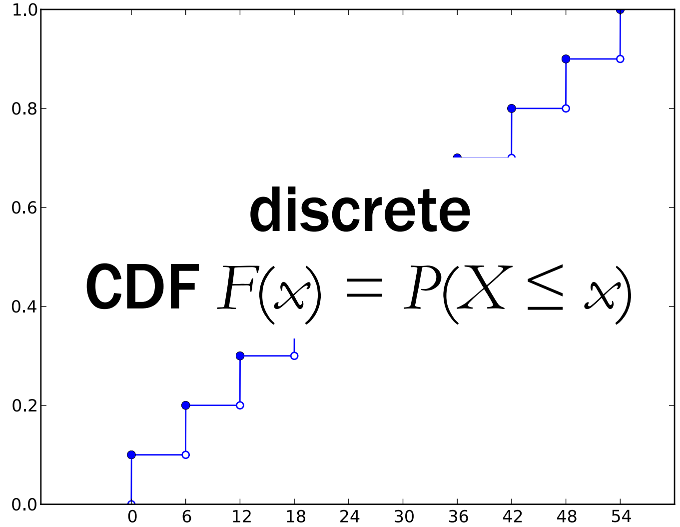
$$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0 & \text{IF } x < 0 \\ 1/60 & \text{IF } x \in [0, 60) \\ 0 & \text{IF } x \geq 60. \end{cases}$$

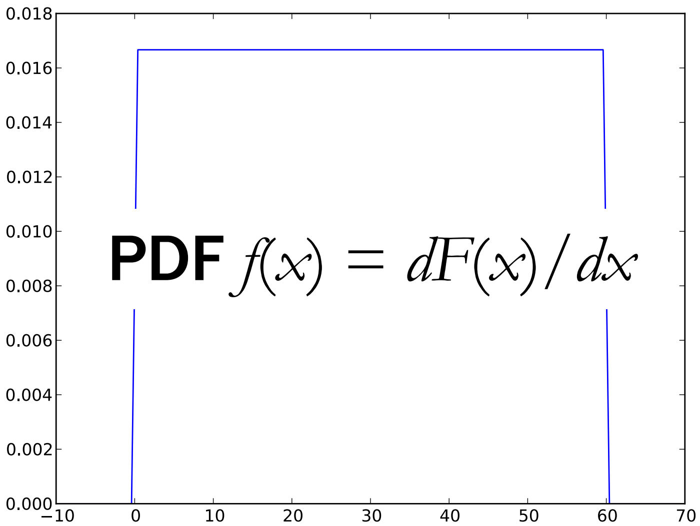
$$\boxed{\int_{-\infty}^{\infty} f(x)\,dx = 1}$$

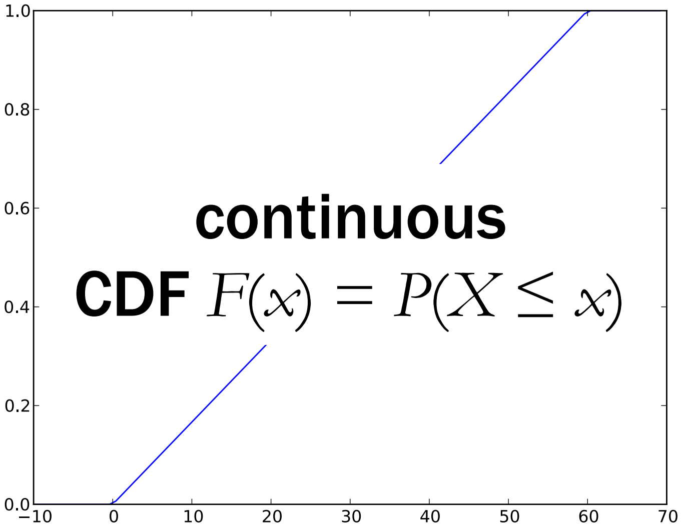


Probability density functions







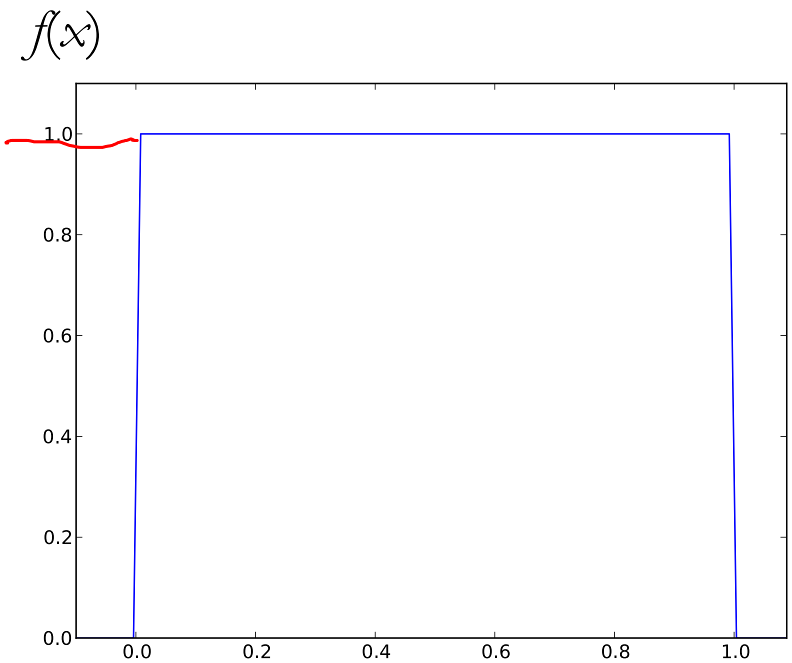


Uniform random variable

The Uniform(0, 1) **PDF is**

$$f(x) = \begin{cases} 1 & \textbf{if } x \in (0, 1) \\ 0 & \textbf{if } x < 0 \textbf{ or } x > 1 \end{cases}$$

$\frac{1}{b-a}$

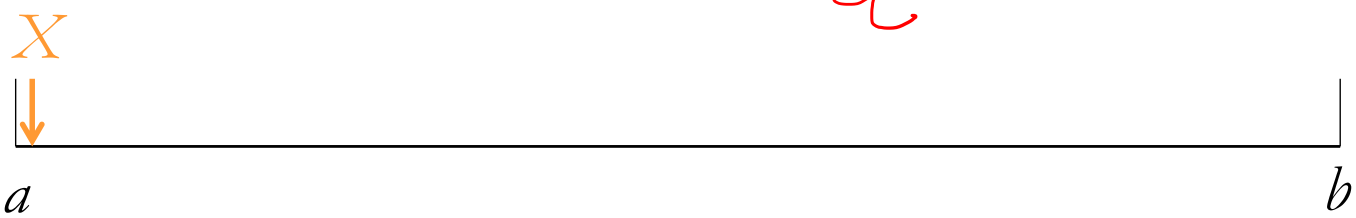


a   b

$a<b$

The Uniform(a, b) **PDF is**

$$f(x) = \begin{cases} 1/(b - a) & \textbf{if } x \in (a, b) \\ 0 & \textbf{if } x < a \textbf{ or } x > b \end{cases}$$



Calculating the CDF

Discrete random variables:

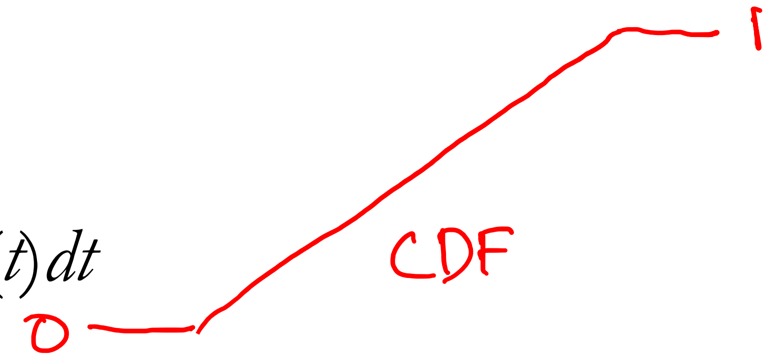
PMF $f(x) = P(X = x)$

CDF $F(x) = P(X \leq x) = \sum_{x \leq t} f(t)$

Continuous random variables:

PDF $f(x) = dF(x)/dx$

CDF $F(x) = P(X \leq x) = \int_{t \leq x} f(t)dt$



A package is to arrive between 12 and 1

What is the probability it arrived by 12.15?

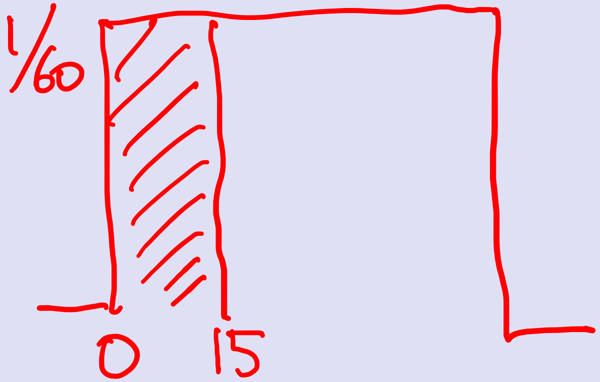
1/4

$X = \text{Uniform}(0, 60)$

$$P(X \leq 15) = \int_{-\infty}^{15} f_X(x)\,dx$$

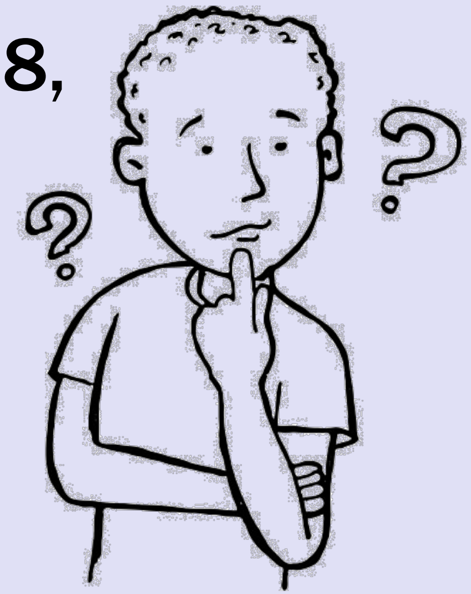
$$= 15 \cdot \frac{1}{60}$$

$$= \frac{1}{4}$$



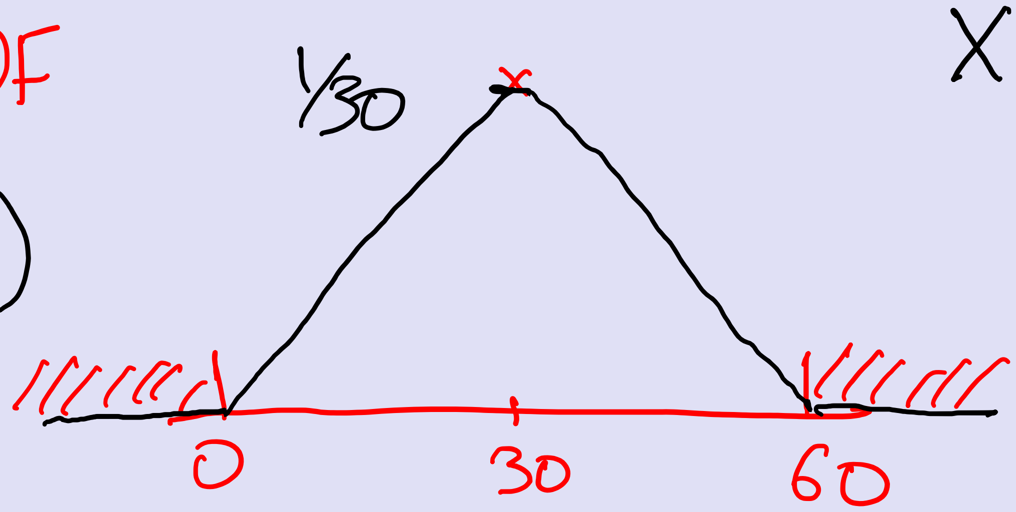
Alice said she'll show up between 7 and 8, probably around 7.30.

It is now 7.30. What is the probability Bob has to wait past 7.45?

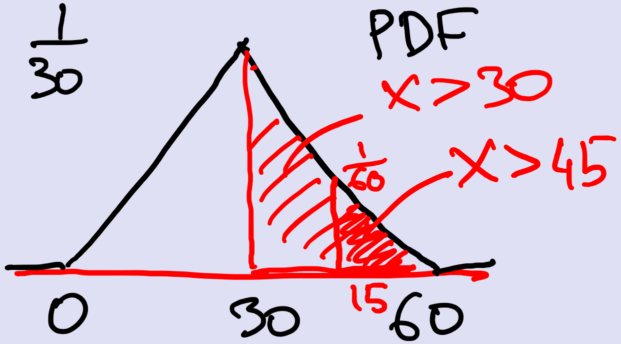


MODEL: ① Uniform(0, 60)

② PDF



$$P(X>45 \mid X>30) = \frac{P(X>45)}{P(X>30)}$$



$$P(X>30)=\frac{1}{2}$$

$$P(X>45)=\frac{15}{60}\cdot\frac{1}{2}=\frac{1}{8}$$

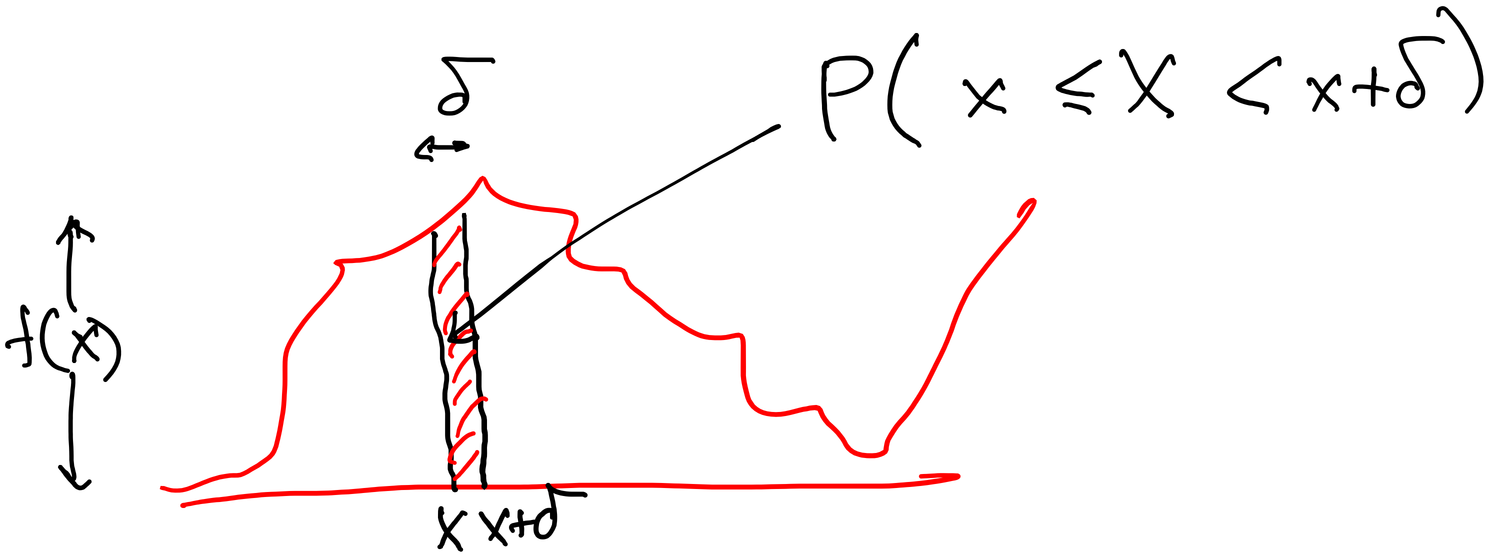
$$P(X>45\mid X>30)=\frac{1/8}{1/2}=1/4$$

Interpretation of the PDF

The PDF value $f(x)\ \delta$ approximates the probability that X in an interval of length δ around x

$$P(x-\delta \le X < x) = f(x)\ \delta + o(\delta)$$

$$P(x \le X < x+\delta) = f(x)\ \delta + o(\delta)$$

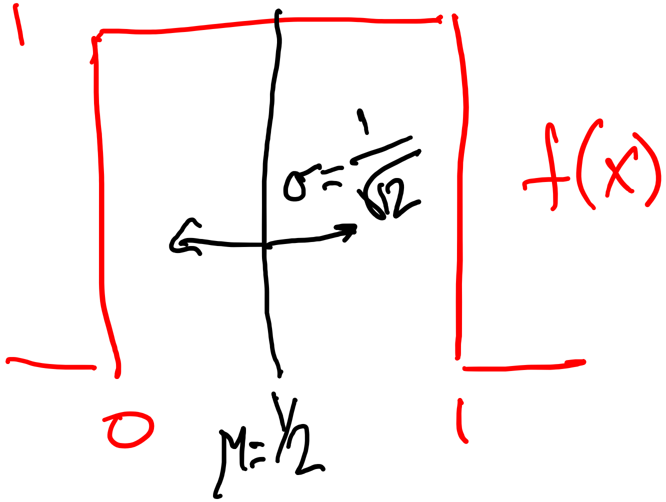


Expectation and variance

	PMF $f(x)$	PDF $f(x)$
$\mathbf{P}(X \leq a)$	$\sum_{x \leq a} f(x)$	$\int_{-\infty}^{a} f(x)\,dx$
$\mathbf{E}[X] = \mu$	$\sum_x x\, f(x)$	$\int_{-\infty}^{\infty} x\, f(x)\,dx$
$\mathbf{E}[X^2]$	$\sum_x x^2 f(x)$	$\int_{-\infty}^{\infty} x^2 f(x)\,dx$
$\mathbf{Var}[X]$	$E[X^2] - E[X]^2$ $= E[(X-\mu)^2]$	$= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

Mean and Variance of Uniform

Uniform(0,1)

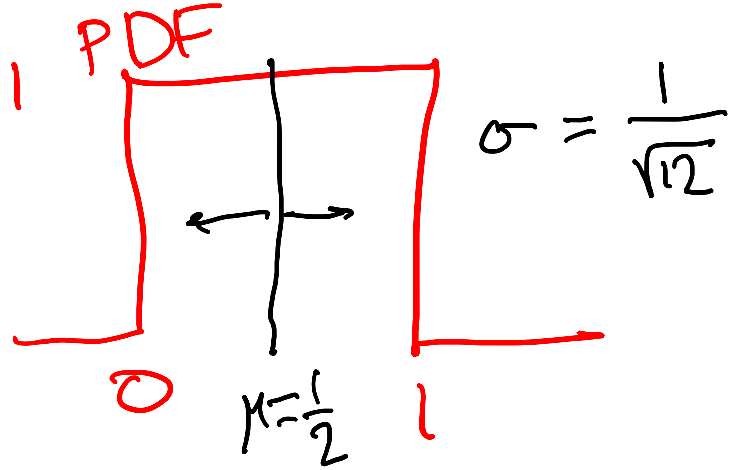


$$E[X] = \int_0^1 x f(x)\,dx$$
$$= \int_0^1 x\,dx$$
$$= \frac{x^2}{2}\Big|_0^1 = \boxed{\frac{1}{2}}$$

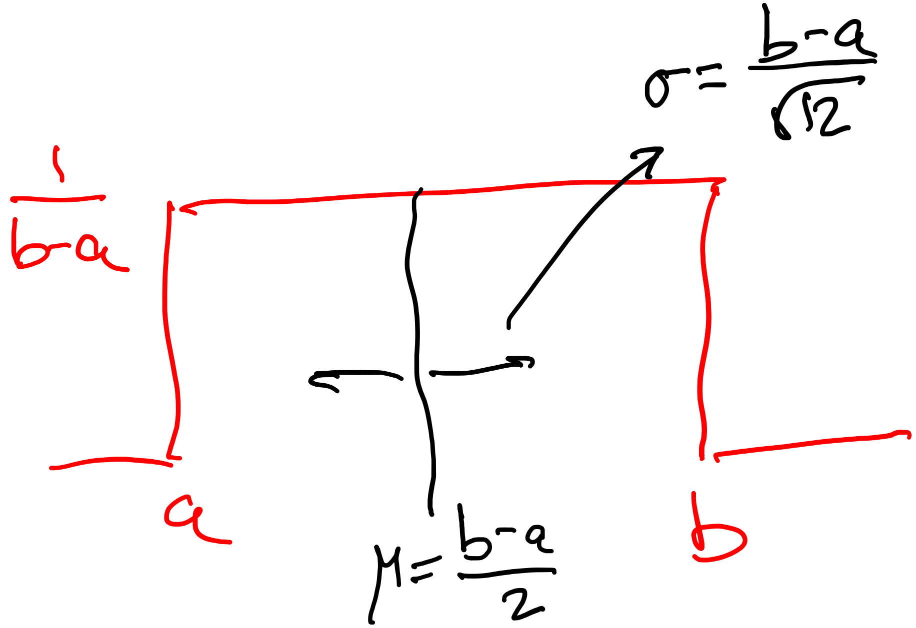
$$E[X^2] = \int_0^1 x^2 f(x)\,dx = \int_0^1 x^2\,dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$$

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \frac{1}{3} - \frac{1}{4} = \boxed{\frac{1}{12}.}$$

Uniform(0,1)



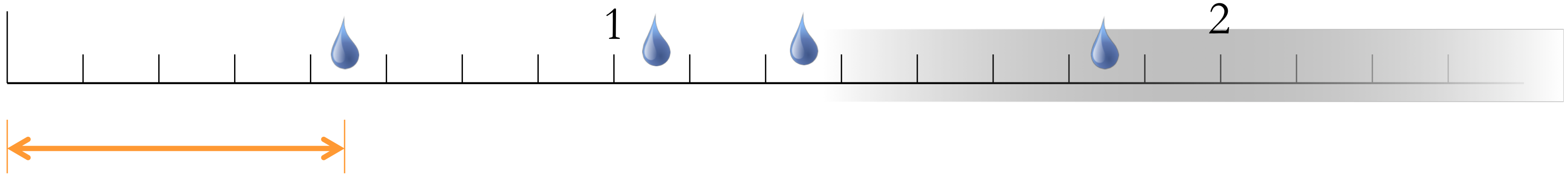
Uniform(a,b)



Raindrops again

Rain is falling on your head at an average speed of λ drops/second.

λ

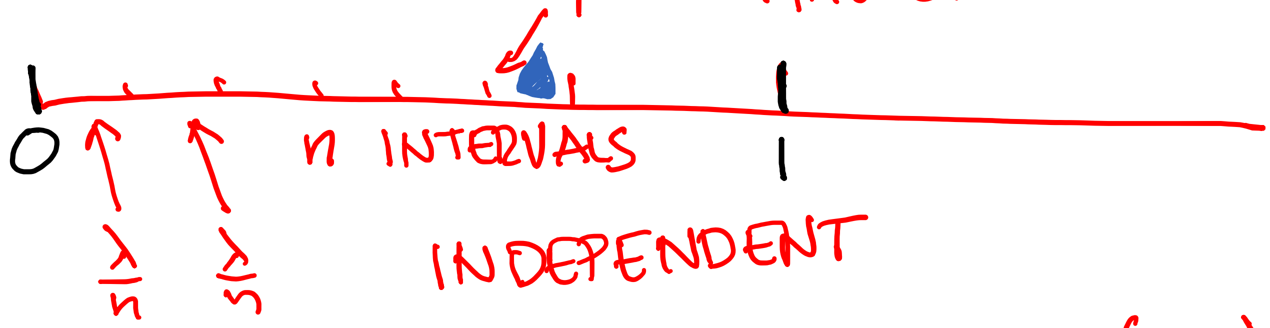


T

How long do we wait until the next drop?

DISCRETE MODEL

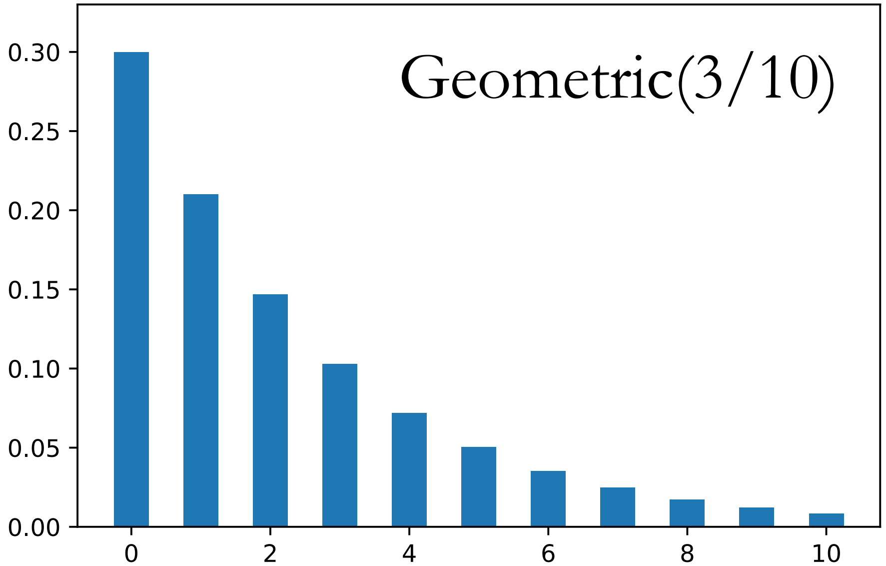
T = TIME OF FIRST DROP

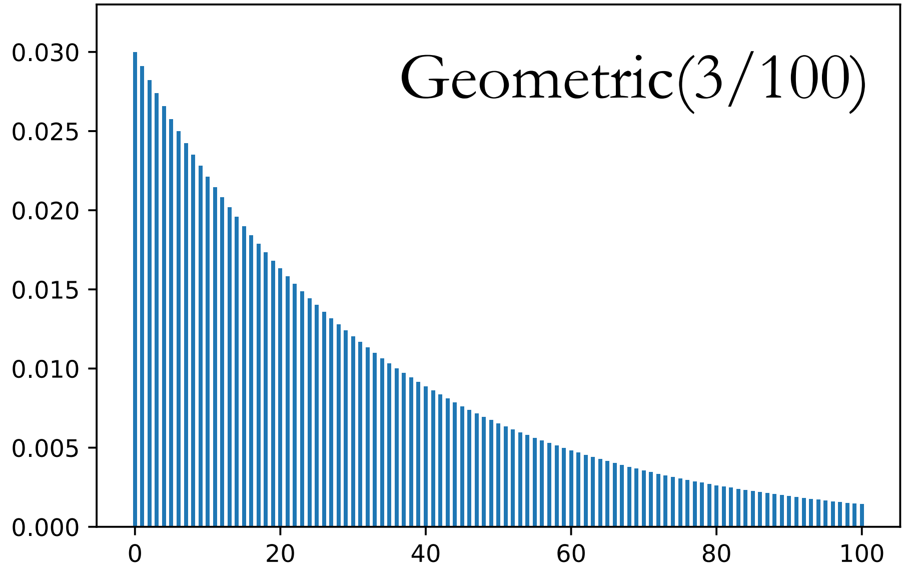


$$P\left(t \le T < t + \frac{1}{n}\right) = P(FF\ldots FS) = \left(1 - \frac{\lambda}{n}\right)^{nt} \cdot \frac{\lambda}{n} \qquad \boxed{\delta = \frac{1}{n}}$$

$$f_T(t) \cdot \delta \approx \left(1 - \frac{\lambda}{n}\right)^{nt} \cdot \lambda \delta$$

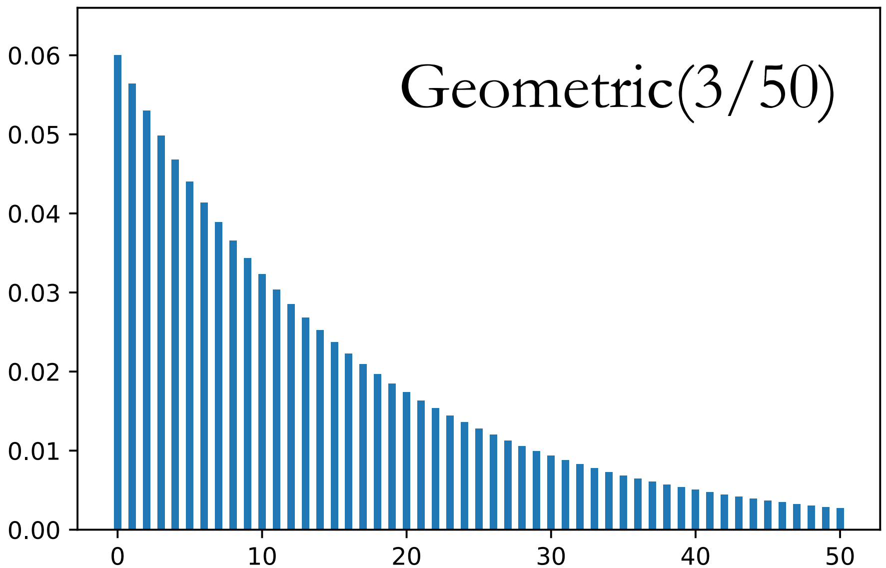
$$f_T(t) = \lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^{nt} \lambda = \lambda e^{-\lambda t}$$

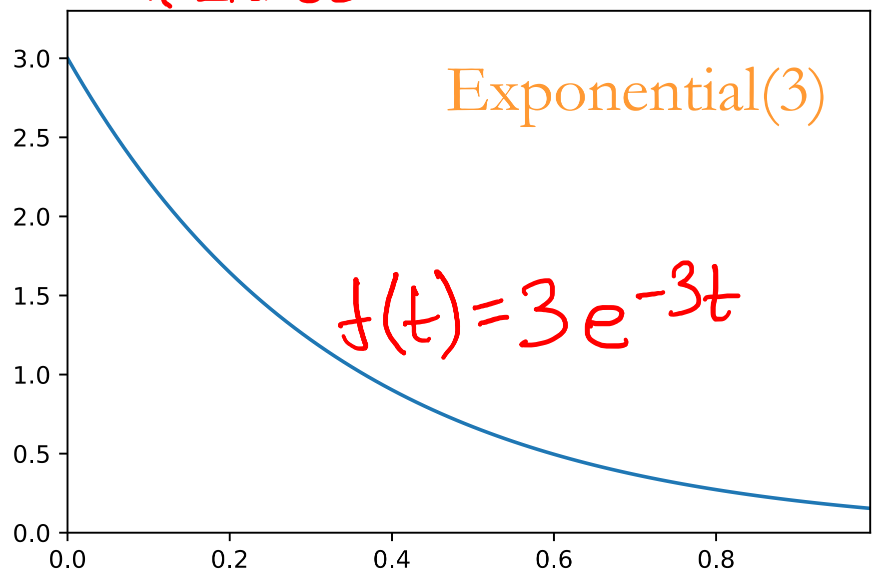




$$\text{Exponential}(\lambda) \approx \frac{1}{n} \cdot \text{Geometric}\left(\frac{\lambda}{n}\right)$$

n LARGE

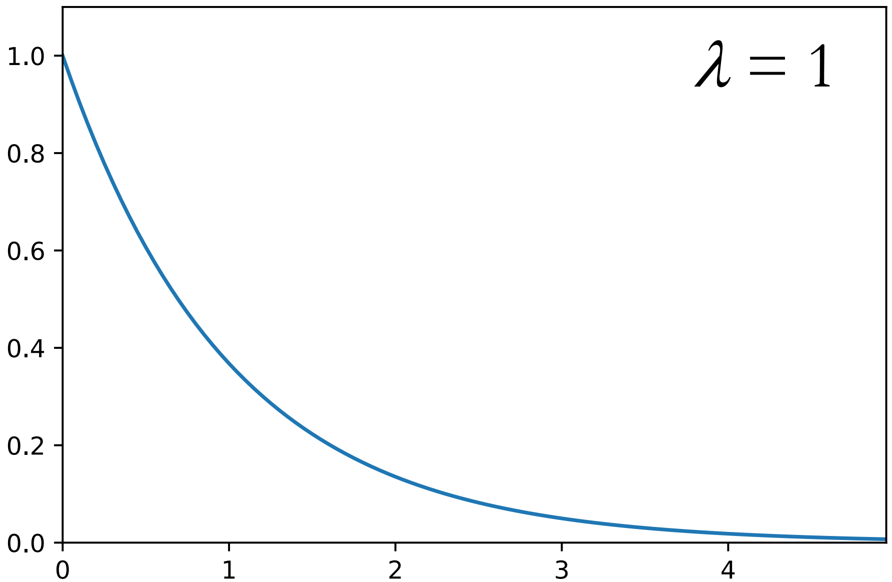




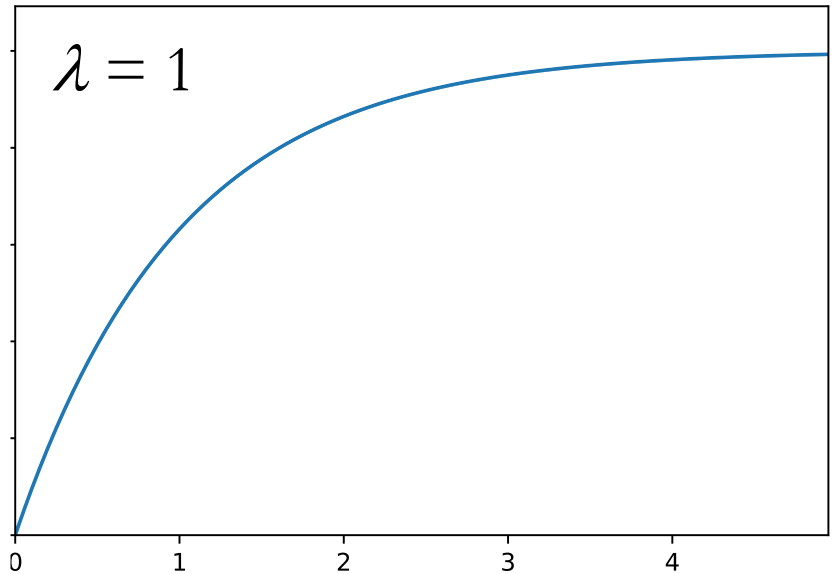
The exponential random variable

The $\text{Exponential}(\lambda)$ **PDF is**

$$f(t) = \begin{cases} \lambda\, e^{-\lambda t} & \textbf{if } x \geq 0 \\ 0 & \textbf{if } x < 0. \end{cases}$$



PDF $f(t)$



CDF $F(t) = P(T \leq t)$

The exponential random variable

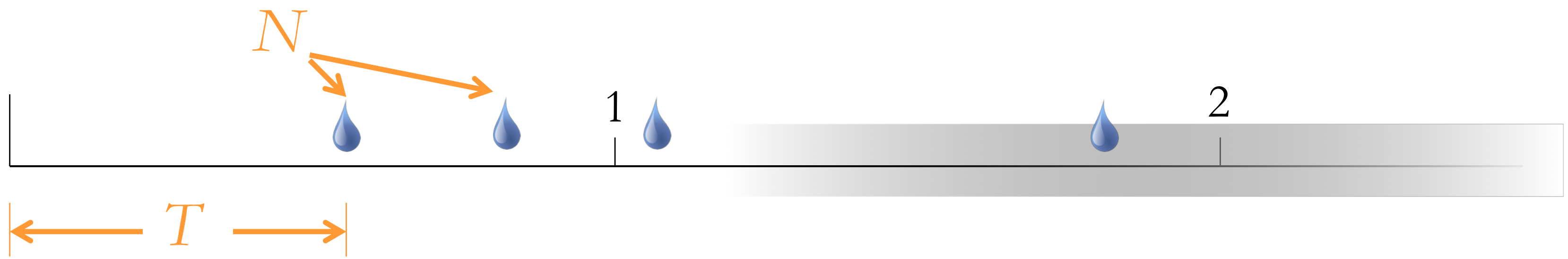
CDF of Exponential(λ): $P(T \le t) = \int_0^t \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_0^t = 1 - e^{-\lambda t}$

$$\text{Exponential}(\lambda) \approx \frac{1}{n}\,\text{Geometric}(p) \qquad p = \frac{\lambda}{n}$$

$$\mathbf{E}[\text{Exponential}(\lambda)] = \lim_{n\to\infty} \frac{1}{n}\cdot\frac{1}{p} = \lim_{n\to\infty} \frac{1}{n}\cdot\frac{n}{\lambda} = \boxed{1/\lambda}$$

$$\mathbf{Var}[\text{Exponential}(\lambda)] = \lim_{n\to\infty} \frac{1}{n^2}\cdot\frac{1-p}{p^2} = \lim_{n\to\infty} \frac{1}{n^2}\cdot\frac{1-\lambda/n}{(\lambda/n)^2} = \boxed{1/\lambda^2}$$

Poisson vs. exponential

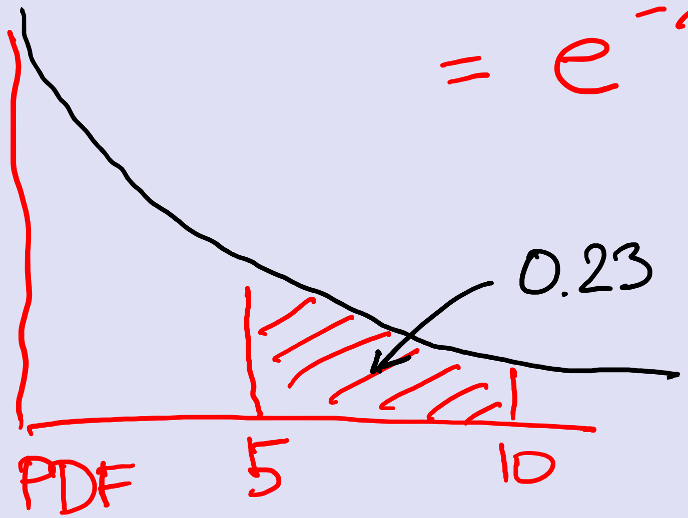


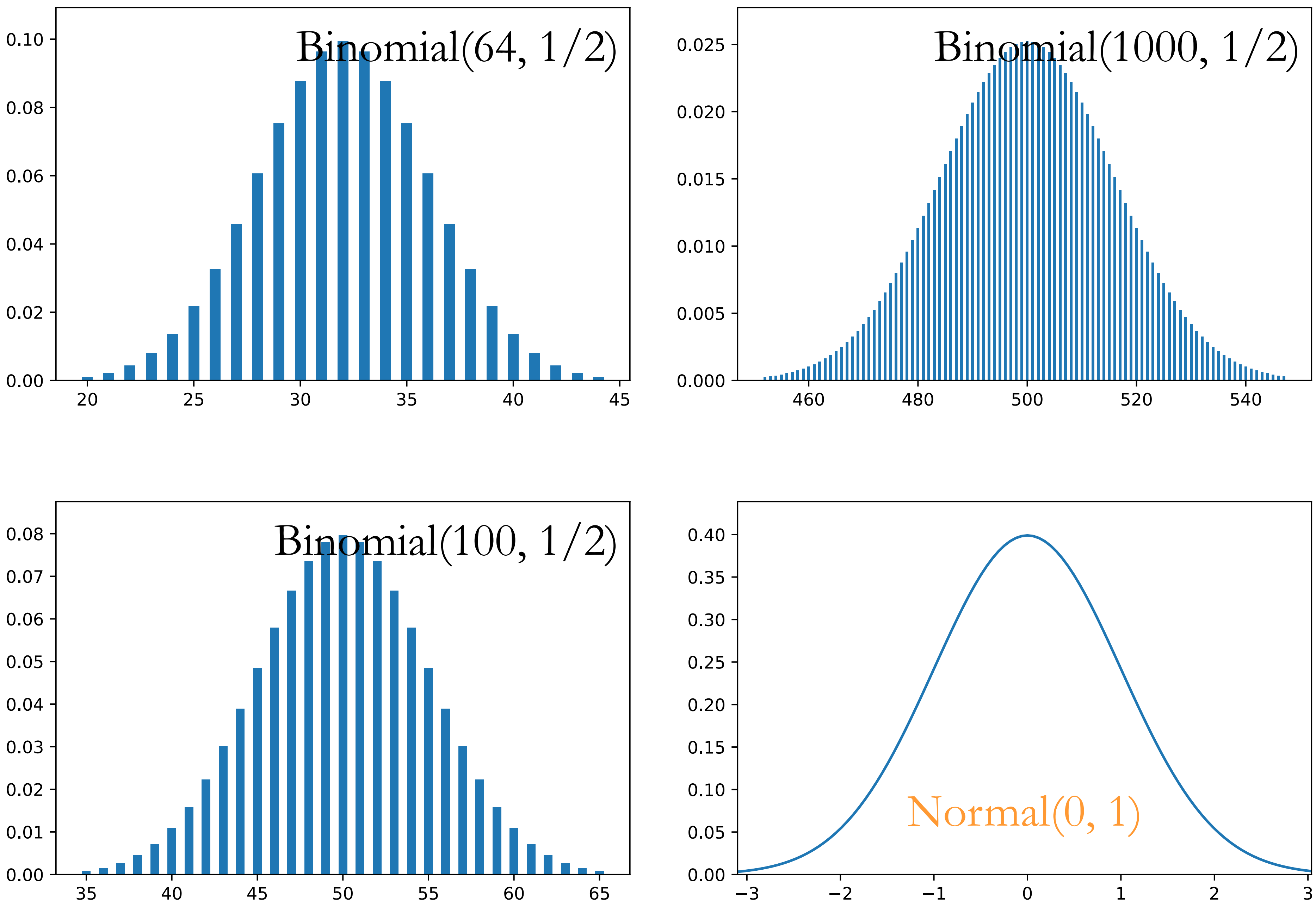
	$\text{Poisson}(\lambda)$	$\text{Exponential}(\lambda)$
description	**number of events within time unit**	**time until first event happens**
expectation	λ	$1/\lambda$
std. deviation	λ	$1/\lambda$

A bus arrives once every 5 minutes. How likely are you to wait 5 to 10 minutes?

Exponential(1/5) RANDOM VARIABLE.

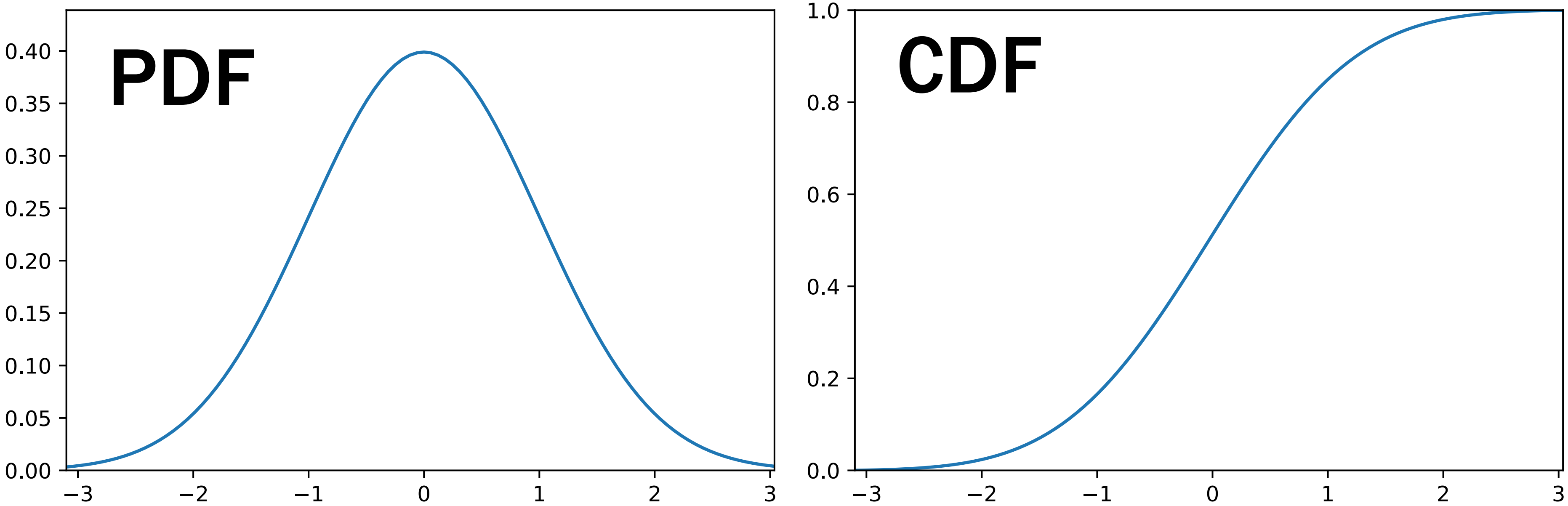
$$P(5 \le T \le 10) = P(T \le 10) - P(T \le 5)$$
$$= \left(1 - e^{-\lambda \cdot 10}\right) - \left(1 - e^{-\lambda \cdot 5}\right)$$
$$= e^{-\lambda \cdot 5} - e^{-\lambda \cdot 10} = \frac{1}{e} - \frac{1}{e^2} \approx 0.23$$





Binomial(64, 1/2)
0.10
0.08
0.06
0.04
0.02
0.00
20
25
30
35
40
45
Binomial(1000, 1/2)
0.025
0.020
0.015
0.010
0.005
0.000
460
480
500
520
540
Binomial(100, 1/2)
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
35
40
45
50
55
60
65
Normal(0, 1)
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−3
−2
−1
0
1
2
3

The Normal(0, 1) random variable



$$f(x) = (2\pi)^{-1/2}\, e^{-x^2/2}$$

$$F(x) = (2\pi)^{-1/2} \int_{t \le x} e^{-t^2/2}\, dt$$

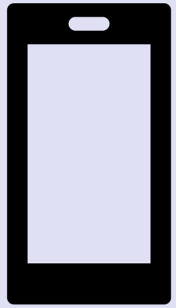
$\mathbf{E}[\text{Normal}(0, 1)] = 0$

$\mathbf{Var}[\text{Normal}(0, 1)] = 1$

Alice

$-1 \longrightarrow -1+N$

$+1 \longrightarrow 1+N$



Bob

$+1 \longrightarrow$ +0.8

+0.3

+1.2

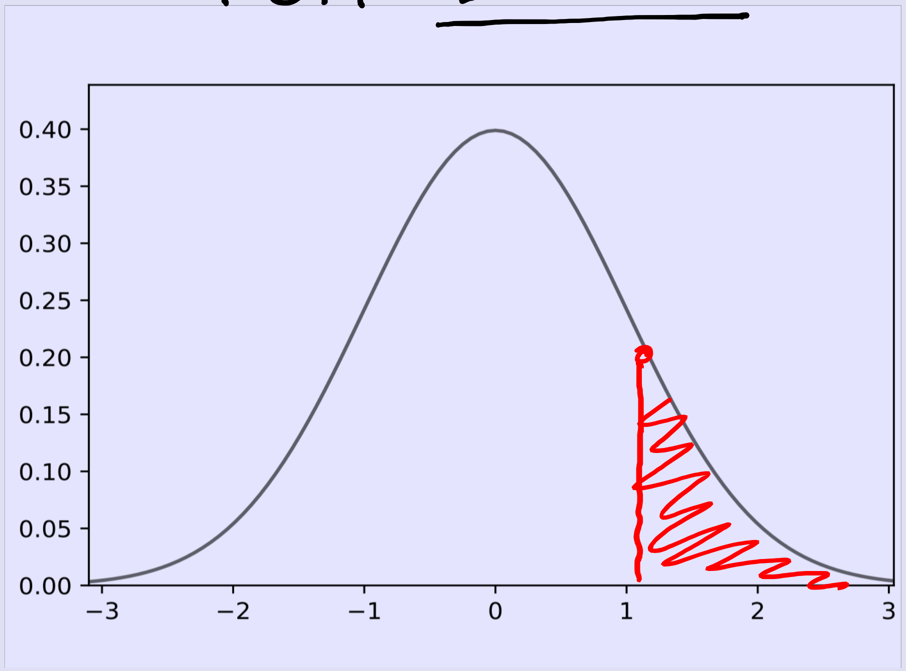
-0.7

$-1 \longrightarrow$ -1.1

+0.1 ERROR!

$$P(\text{ERROR}) = \int_1^{\infty} f_N(t)\,dt$$

$$\approx 0.1587$$



The Normal(μ, σ) random variable

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

$E[\text{Normal}(\mu, \sigma)] = \mu$

$\text{Var}[\text{Normal}(\mu, \sigma)] = \sigma^2$