Consider i.i.d. (independent, identically distributed) random vars $X_1, X_2, X_3, \ldots$

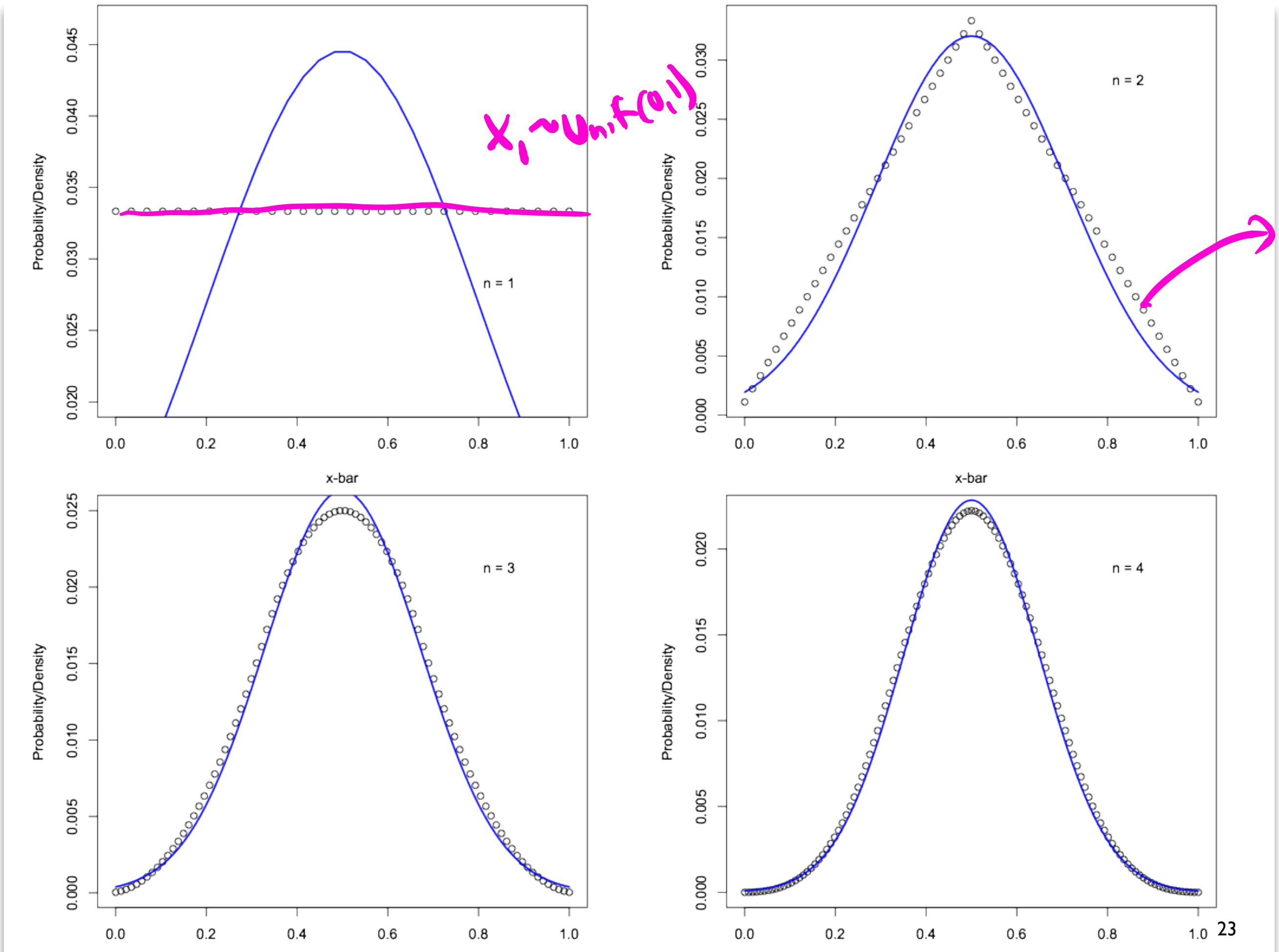
X_i has $\mu = E[X_i]$ and $\sigma^2 = Var[X_i]$

As $n \to \infty$,

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0,1)$$

Restated: As $n \to \infty$,

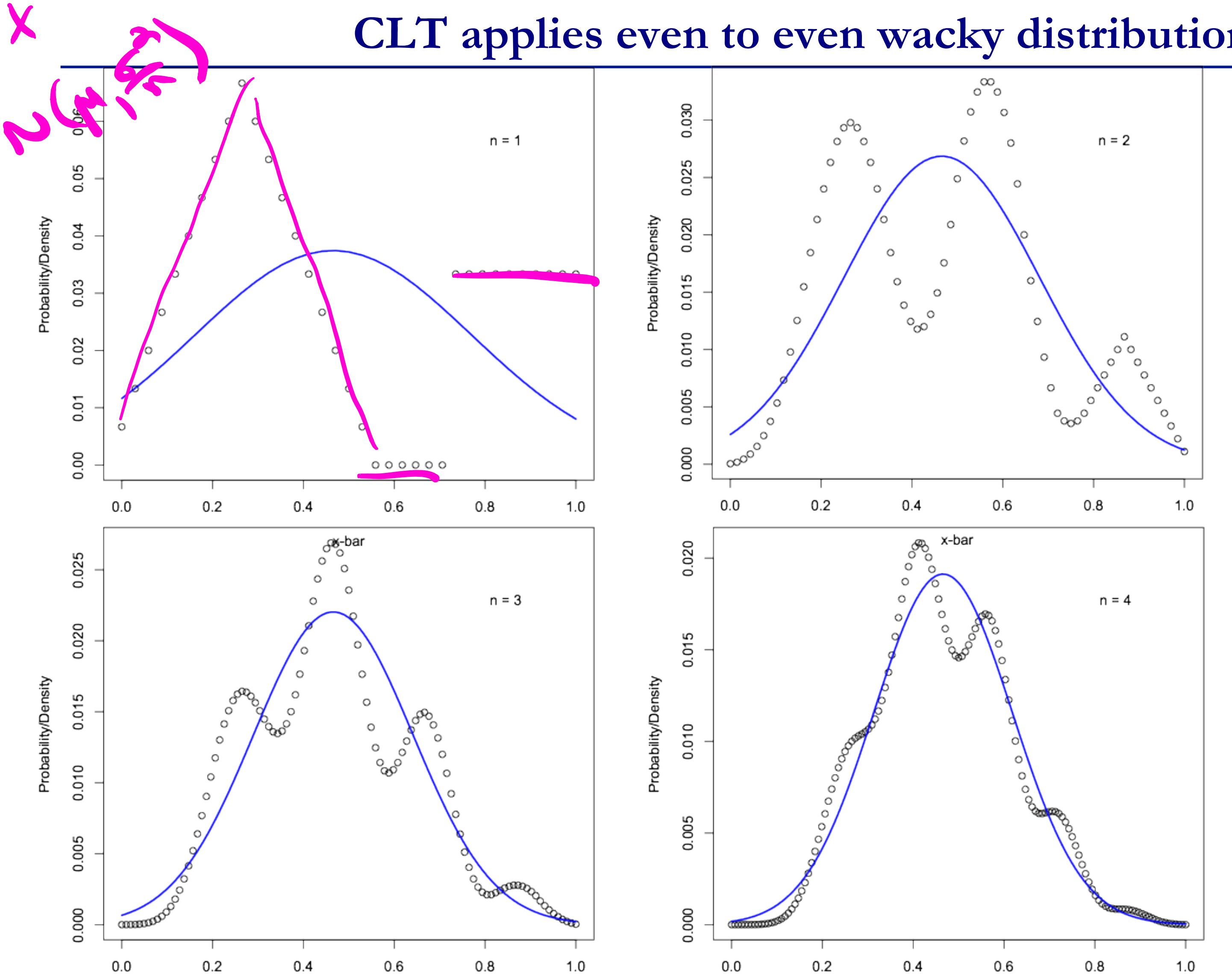
$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to N\left(\mu, \frac{\sigma^2}{n}\right)$$

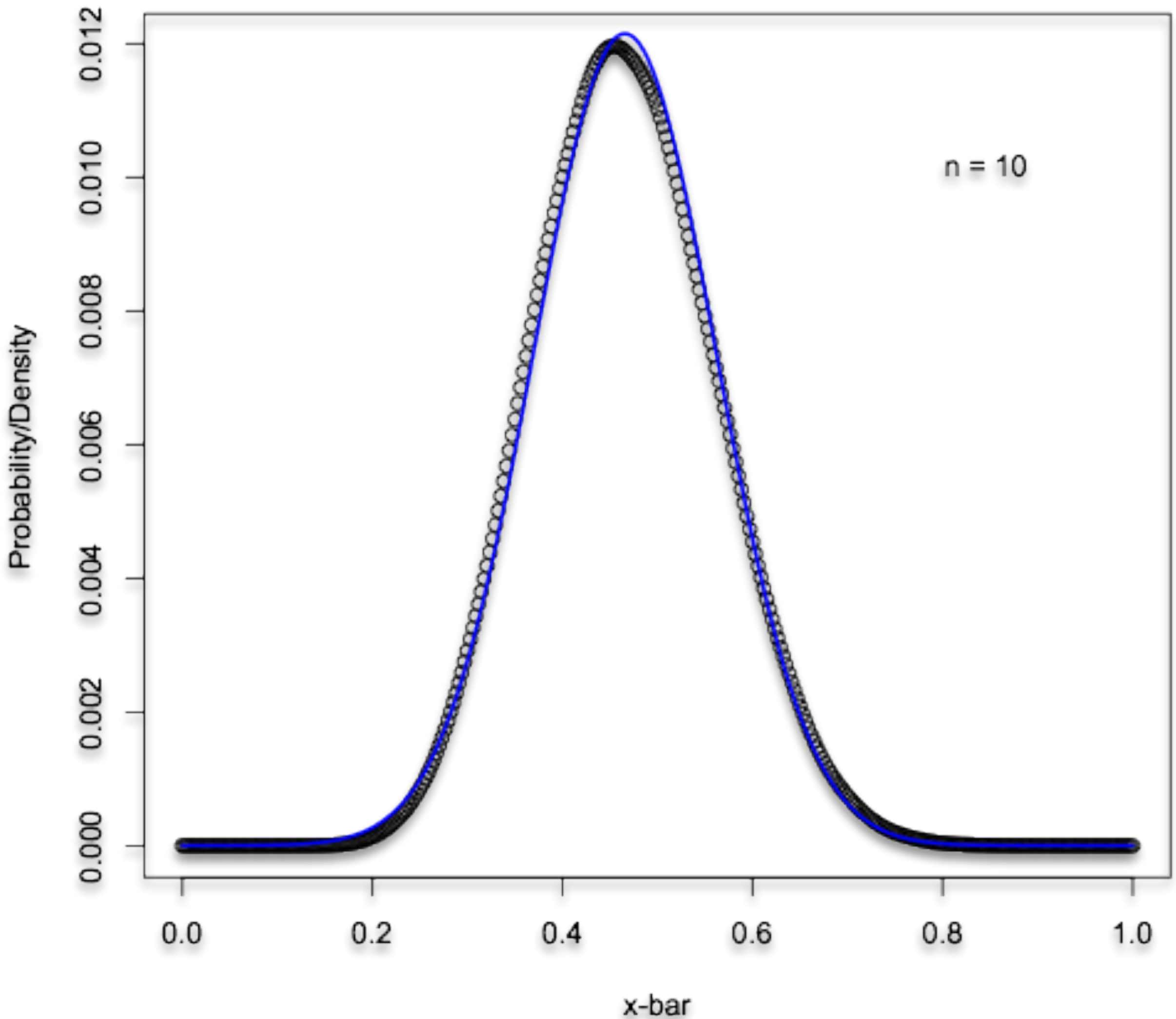


$X_1 \sim \mathrm{Unif}(0,1)$

$\frac{X_1 + X_2}{2}$

CLT applies even to even wacky distributions



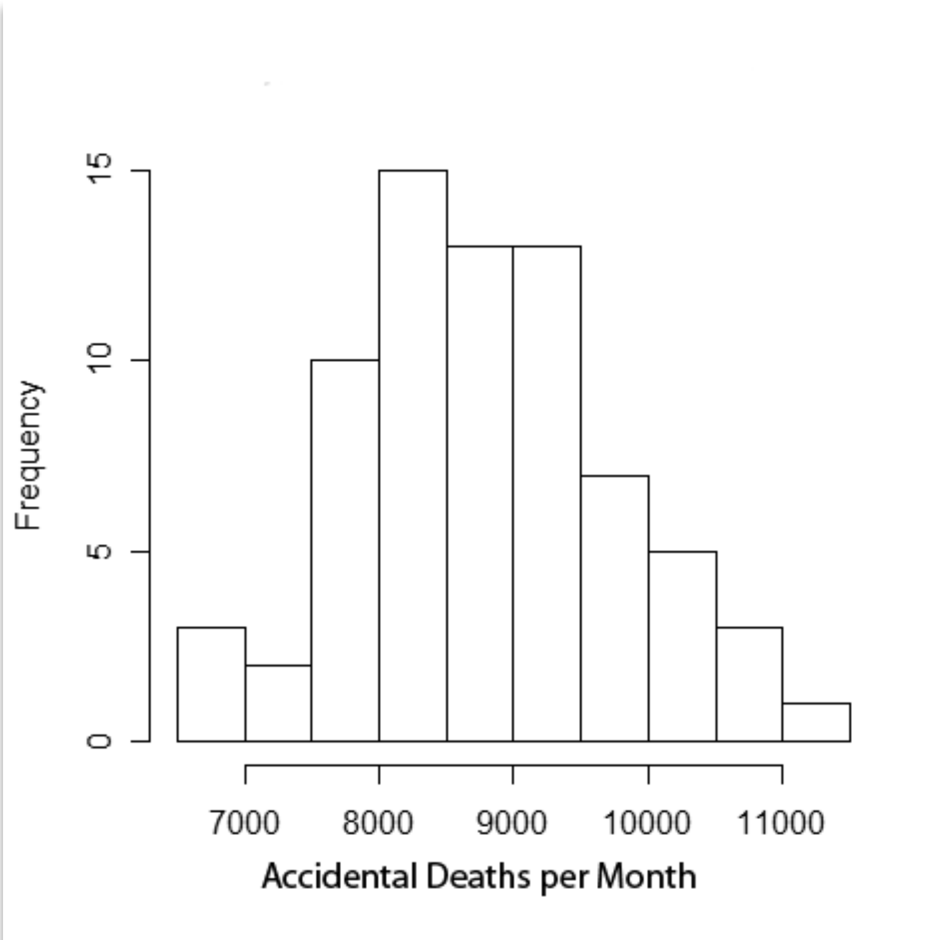


n = 10
Probability/Density
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.0
0.2
0.4
0.6
0.8
1.0
x-bar

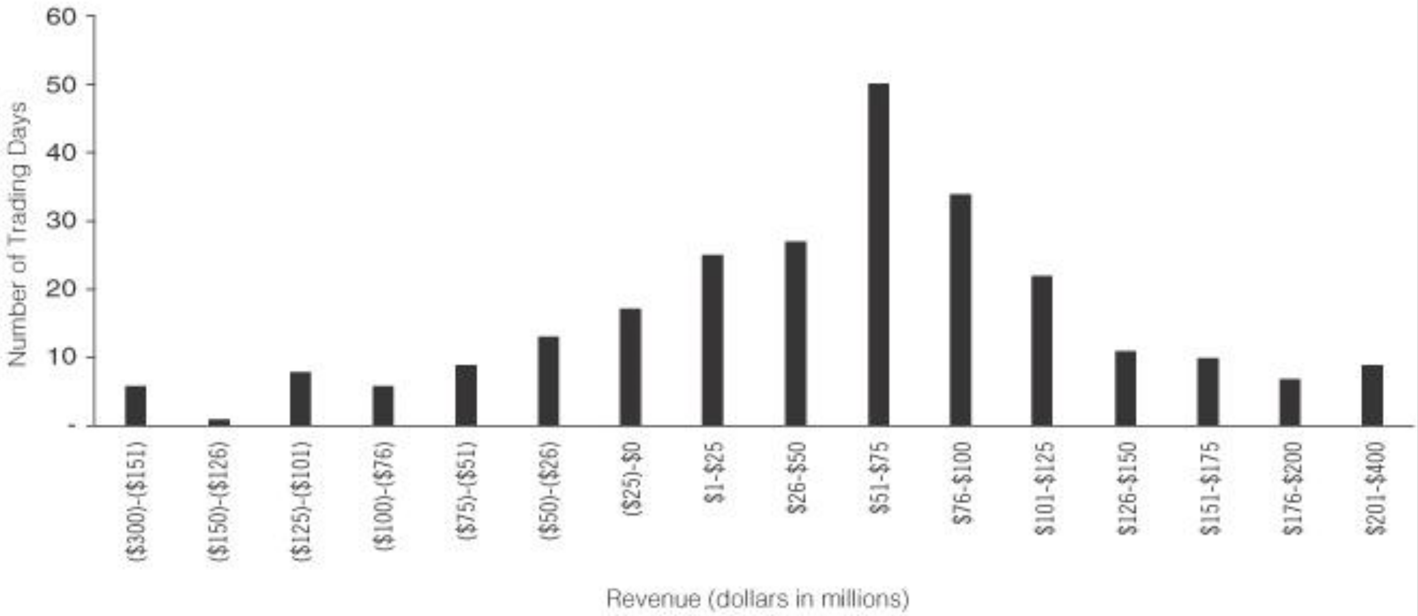
CLT in the real world

CLT is the reason many things appear normally distributed
Many quantities = sums of (roughly) independent random vars

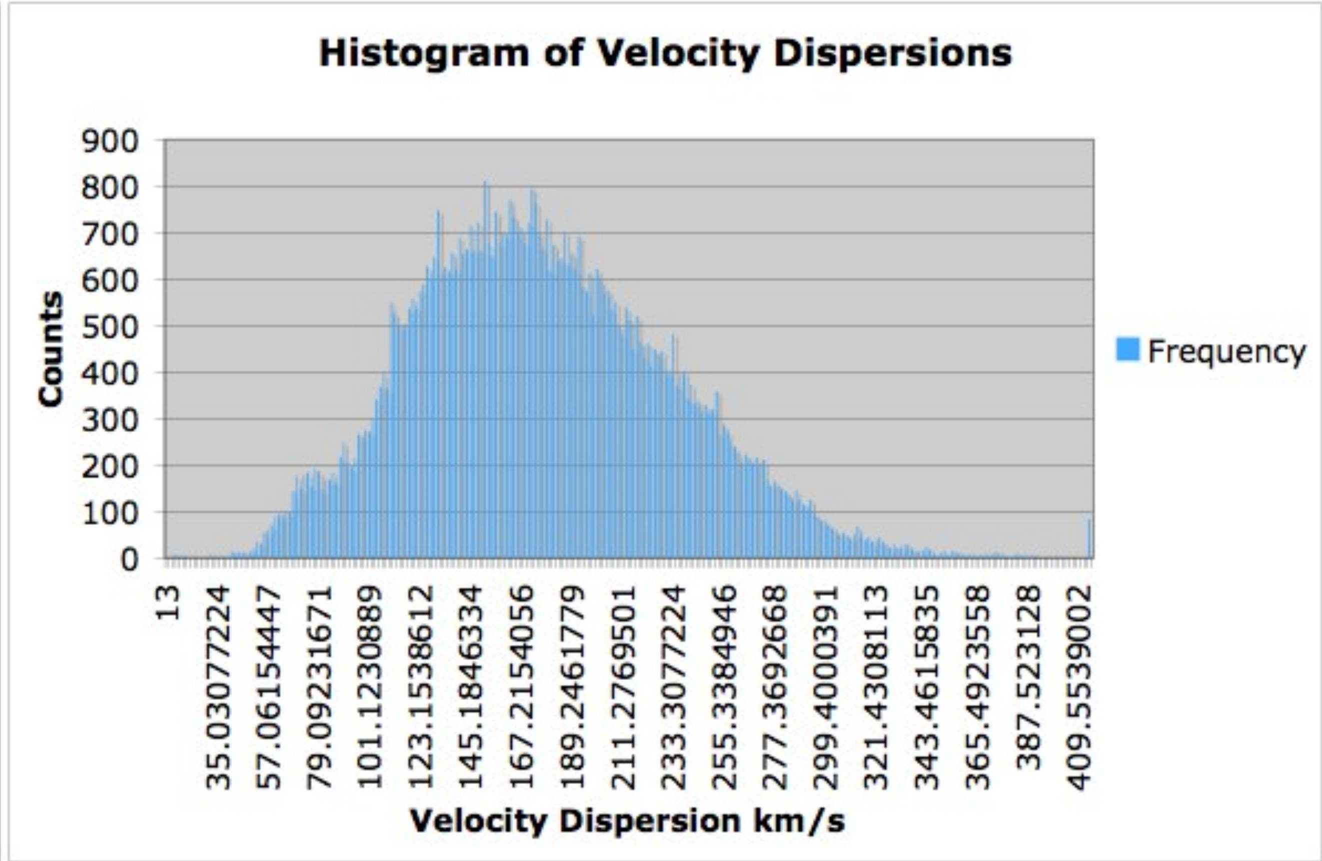
Exam scores: sums of individual problems
People's heights: sum of many genetic & environmental factors
Measurements: sums of various small instrument errors
...


15
10
5
0
Frequency
7000
8000
9000
10000
11000
Accidental Deaths per Month

Histogram of Daily Trading-Related Revenue* — Twelve Months Ended December 31, 2007



*Excludes daily profits and losses in the ABS CDO market, including recent subprime-related losses.


Histogram of Velocity Dispersions
Counts
900
800
700
600
500
400
300
200
100
0
13
35.030772224
57.061544447
79.092231671
101.12308889
123.15386112
145.18463334
167.21540556
189.24617779
211.27695001
233.30772224
255.33849446
277.36926668
299.40003891
321.43081113
343.46158335
365.49235558
387.5231128
409.55390002
Velocity Dispersion km/s
Frequency

Consider i.i.d. (independent, identically distributed) random vars $X_1, X_2, X_3, \ldots$

X_i has $\mu = E[X_i]$ and $\sigma^2 = Var[X_i]$

As $n \to \infty$,

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0,1)$$

Restated: As $n \to \infty$,

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to N\left(\mu, \frac{\sigma^2}{n}\right)$$

Discrete	Continuous
$p_{X,Y}(x,y) = P(X = x, Y = y)$	$f_{X,Y}(x,y) \neq P(X = x, Y = y)$
$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s) ds dt$
$\sum_{x} \sum_{y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$
$p_X(x) = \sum_{y} p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$
$E[g(X,Y)] = \sum_{x} \sum_{y} g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$

Independence	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$