	Discrete: pmf $p(\cdot)$	Continuous: pdf $f(\cdot)$
Support Set	Countable set of values	Uncountable set of values
Probabilities	$p(x) = P(X = x)$	$f(x) \neq P(X = x) = 0$ for all x $P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$
Joint	$p_{XY}(x, y) = \mathbb{P}(X = x, Y = y)$	$\mathbb{P}(X \in [a, b], Y \in [c, d]) = \int_a^b \int_c^d f_{XY}(x, y)\, dx\, dy$
Marginal	$p_X(x) = \sum_y p_{XY}(x, y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy$
Conditional	$p_{X\|Y}(x\|y) = p_{XY}(x, y)/p_Y(y)$	$f_{X\|Y}(x\|y) = f_{XY}(x, y)/f_Y(y)$
Independence	$p_{XY}(x, y) = p_X(x)p_Y(y)$	$f_{XY}(x, y) = f_X(x)f_Y(y)$
Expected Value	$\mu_X = \mathbb{E}[X] = \sum_x xp(x)$ $\mathbb{E}[g(X)] = \sum_x g(x)p(x)$ $\mathbb{E}[g(X, Y)] = \sum_x \sum_y g(x, y)p(x, y)$	$\mu_X = \mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)\, dx$ $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\, dx$ $\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f_{XY}(x, y)\, dx\, dy$

Table 1: Differences between Discrete and Continous Random Variables

Probability Mass Function $p(x)$	Probability Density Function $f(x)$
Discrete Random Variables	Continuous Random Variables
$p(x) = P(X = x)$	$f(x) \neq P(X = x) = 0$
$p(x) \geq 0$	$f(x) \geq 0$
$p(x) \leq 1$	$f(x)$ *can* be greater than one!
$\sum_x p(x) = 1$	$\int_{-\infty}^{\infty} f(x)\, dx = 1$
$F(x_0) = \sum_{x \leq x_0} p(x)$	$F(x_0) = \int_{-\infty}^{x_0} f(t)\, dt$

Table 2: Probability mass function (pmf) versus probability density function.

Definition of R.V.	$X\colon S \to \mathbb{R}$ (RV is a fixed function from sample space to reals)
Support Set	Collection of all possible realizations of a RV
CDF	$F(x_0) = P(X \leq x_0)$
Expectation of a Function	In general, $\mathbb{E}[g(X)] \neq g\left(\mathbb{E}[X]\right)$
Linearity of Expectation	$\mathbb{E}[a+X] = a + \mathbb{E}[X], \quad \mathbb{E}[bX] = b\mathbb{E}[X], \quad \mathbb{E}[X_1 + \ldots + X_k] = \mathbb{E}[X_1] + \ldots \mathbb{E}[X_k]$
Variance	$\sigma_X^2 \equiv \text{Var}(X) \equiv \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}\left[X(X-\mu_X)\right]$
Standard Deviation	$\sigma_X = \sqrt{\sigma_X^2}$
Var. of Linear Combination	$\text{Var}(a+X) = \text{Var}(X), \quad \text{Var}(bX) = b^2\text{Var}(X), \quad \text{Var}(aX + bY + c) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X,Y)$
	$X_1, \ldots, X_k$ are uncorrelated $\Rightarrow \text{Var}(X_1 + \ldots + X_k) = \text{Var}(X_1) + \ldots \text{Var}(X_k)$
Covariance	$\sigma_{XY} \equiv \text{Cov}(X,Y) \equiv \mathbb{E}\left[(X - \mathbb{E}[X])\,(Y - \mathbb{E}[Y])\right] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}\left[X(Y - \mu_Y)\right] = \mathbb{E}\left[(X - \mu_X)Y\right]$
Correlation	$\rho_{XY} = \text{Corr}(X,Y) = \sigma_{XY}/(\sigma_X\sigma_Y)$
Covariance and Independence	X,Y independent $\Rightarrow \text{Cov}(X,Y) = 0$ but $\text{Cov}(X,Y) = 0 \not\Rightarrow X,Y$ independent
Functions and Independence	X,Y independent $\Rightarrow g(X), h(Y)$ independent
Bilinearity of Covariance	$\text{Cov}(a+X,Y) = \text{Cov}(X,a+Y) = \text{Cov}(X,Y), \quad \text{Cov}(bX,Y) = \text{Cov}(X,bY) = b\text{Cov}(X,Y)$
	$\text{Cov}(X,Y+Z) = \text{Cov}(X,Y) + \text{Cov}(X,Z)$ and $\text{Cov}(X+Z,Y) = \text{Cov}(X,Y) + \text{Cov}(Z,Y)$
Linearity of Conditional $\mathbb{E}$	$\mathbb{E}[a+Y\|X] = a + \mathbb{E}[Y\|X], \quad \mathbb{E}[bY\|X] = b\mathbb{E}[Y\|X], \quad \mathbb{E}[X_1 + \cdots + X_k\|Z] = \mathbb{E}[X_1\|Z] + \cdots + \mathbb{E}[X_k\|Z]$
Taking Out What is Known	$\mathbb{E}[g(X)Y\|X] = g(X)\mathbb{E}[Y\|X]$
Law of Iterated Expectations	$\mathbb{E}[Y] = \mathbb{E}\left[\mathbb{E}(Y\|X)\right]$
Conditional Variance	$\text{Var}(Y\|X) \equiv \mathbb{E}\left\{(Y - \mathbb{E}[Y\|X])^2 \middle\| X\right\} = \mathbb{E}[Y^2\|X] - (\mathbb{E}[Y\|X])^2$
Law of Total Variance	$\text{Var}(Y) = \mathbb{E}\left[\text{Var}(Y\|X)\right] + \text{Var}\left(\mathbb{E}[Y\|X]\right)$

Table 3: Essential facts that hold for *all* random variables, continuous or discrete: X, Y, Z and $X_1, \ldots, X_k$ are random variables; a, b, c, d are constants; μ, σ, ρ are parameters; and $g(\cdot)$, $h(\cdot)$ are functions.

	Sample Statistic	Population Parameter	Population Parameter
Setup	Sample of size $n < N$ from a popn.	Population viewed as list of N objects	Population viewed as a RV
Mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\mu_X = \frac{1}{N}\sum_{i=1}^{N} x_i$	Discrete $\mu_X = \sum_{x} xp(x)$ Continuous $\mu_X = \int_{-\infty}^{\infty} xf(x)\, dx$
Variance	$s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	$\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_X)^2$	$\sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$
Std. Dev.	$s_X = \sqrt{s_X^2}$	$\sigma_X = \sqrt{\sigma_x^2}$	$\sigma_X = \sqrt{\sigma_x^2}$
Covariance	$s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$	$\sigma_{XY} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_X)(y_i - \mu_Y)$	$\sigma_{XY} = E\left[(X - \mu_X)(Y - \mu_Y)\right]$
Correlation	$r_{XY} = s_{XY}/(s_X s_Y)$	$\rho_{XY} = \sigma_{XY}/(\sigma_X \sigma_Y)$	$\rho_{XY} = \sigma_{XY}/(\sigma_X \sigma_Y)$