Math 5760/6890: Introduction to Mathematical Finance
Review: probability

Akil Narayan[1]

[1] Department of Mathematics, and Scientific Computing and Imaging (SCI) Institute
University of Utah

September 7, 2023

www.sci.utah.edu

Our status in the course

We've discussed the basics of finance and investing – concepts of interest and present value.

A more advanced understanding of pricing and policies requires some math:

- linear algebra
- differential equations
- **probability**

These topics are prerequisites for this course!

Events

Probability is a language about potential outcomes; these potential outcomes are called *events*.

A foundational concept is the *event space*, which is the set of all possible outcomes.

Events

Probability is a language about potential outcomes; these potential outcomes are called *events*.

A foundational concept is the *event space*, which is the set of all possible outcomes.

Example

I roll a 6-sided fair die. The possible events are:

- Face 1 is on top
- $\vdots$
- Face 6 is on top

Note that the *numbers* 1 through 6 are *not* the events.

Events

Probability is a language about potential outcomes; these potential outcomes are called *events*.

A foundational concept is the *event space*, which is the set of all possible outcomes.

Example

I roll a 6-sided fair die. The possible events are:

- Face 1 is on top
- ⋮
- Face 6 is on top

Note that the *numbers* 1 through 6 are *not* the events.

Another example: I play paper-rock-scissors, and I'm concerned with which object I play (ignoring my opponent). The possible events are:

- I play paper
- I play scissors
- I play rock

Random variables

We typically deal with numeric values assigned to events. These assignments are called *random variables*.[1]

Typically, the assignment of events to numerical values is somewhat straightforward.

[1] More abstractly, random variables are "well-defined" functions that map events to real numbers.

Random variables

We typically deal with numeric values assigned to events. These assignments are called *random variables*.[1]

Typically, the assignment of events to numerical values is somewhat straightforward.

Example

I roll a 6-sided fair die. It's quite sensible for me to define a random variable X to denote the label of the side that comes up:

$$\underbrace{\text{Face 3 is on top}}_{\text{event}} \longrightarrow \underbrace{X = 3}_{\text{Random variable assignment}}$$

The set of possible values of the random variable (here X) is $\{1, 2, 3, 4, 5, 6\}$.

[1] More abstractly, random variables are "well-defined" functions that map events to real numbers.

Random variables

We typically deal with numeric values assigned to events. These assignments are called *random variables*.[1]

Typically, the assignment of events to numerical values is somewhat straightforward.

Example

I roll a 6-sided fair die. It's quite sensible for me to define a random variable X to denote the label of the side that comes up:

$$\underbrace{\text{Face 3 is on top}}_{\text{event}} \longrightarrow \underbrace{X = 3}_{\text{Random variable assignment}}$$

The set of possible values of the random variable (here X) is $\{1, 2, 3, 4, 5, 6\}$.

Second example: I play paper-rock-scissors. Here is one random variable definition:

$$\begin{aligned} \text{I play paper} &\longrightarrow Y = 1 \\ \text{I play scissors} &\longrightarrow Y = 2 \\ \text{I play rock} &\longrightarrow Y = 3. \end{aligned}$$

But $Z = 4 - Y$ is a perfectly acceptable, alternative encoding of outcomes. Both Y and Z are sensible random variables, and there is no reason to prefer one to another without further context.

[1] More abstractly, random variables are "well-defined" functions that map events to real numbers.

Probability distributions

The final ingredient we require is a *distribution* on outcomes, that is a definition of likelihood that certain events happen.

We call these likelihoods *probabilities*, and they are always non-negative numbers between 0 and 1, and the sum of probabilities over all outcomes is always 1.

The probability of an event is typically denoted $P(\text{event})$ or $\Pr(\text{event})$.

Probability distributions

The final ingredient we require is a *distribution* on outcomes, that is a definition of likelihood that certain events happen.

We call these likelihoods *probabilities*, and they are always non-negative numbers between 0 and 1, and the sum of probabilities over all outcomes is always 1.

The probability of an event is typically denoted $P(\text{event})$ or $\Pr(\text{event})$.

Example

I roll a 6-sided fair die, and assign the following distribution on outcomes:

$$P(\text{Face 1 is on top}) = \frac{1}{6}, \quad \cdots \quad P(\text{Face 6 is on top}) = \frac{1}{6}.$$

Probability distributions

The final ingredient we require is a *distribution* on outcomes, that is a definition of likelihood that certain events happen.

We call these likelihoods *probabilities*, and they are always non-negative numbers between 0 and 1, and the sum of probabilities over all outcomes is always 1.

The probability of an event is typically denoted $P(\text{event})$ or $\Pr(\text{event})$.

Example

I roll a 6-sided fair die, and assign the following distribution on outcomes:

$$P(\text{Face 1 is on top}) = \frac{1}{6}, \quad \cdots \quad P(\text{Face 6 is on top}) = \frac{1}{6}.$$

Probabilities could be defined only on "coarser" events: I roll a 6-sided die (not necessarily fair), which has the following distribution:

$$P(\text{An even-number-labeled face is on top}) = \frac{1}{2},$$
$$P(\text{An odd-number-labeled face is on top}) = \frac{1}{2}.$$

Note that none of this is directly related to random variables! These are purely properties on the space of outcomes.

Probability mass functions

The examples we've seen are examples where the random variable takes on a discrete (in particular finite) number of values.

For such discrete random variables, a standard practice is to translate probabilities on outcomes into probabilities on variable values:

$$p_X(x) := P\left(X = x\right).$$

p_X is called the *(probability) mass function* for X, and maps elements from the set of values of X to the set of numbers $[0, 1]$.

Probability mass functions

The examples we've seen are examples where the random variable takes on a discrete (in particular finite) number of values.

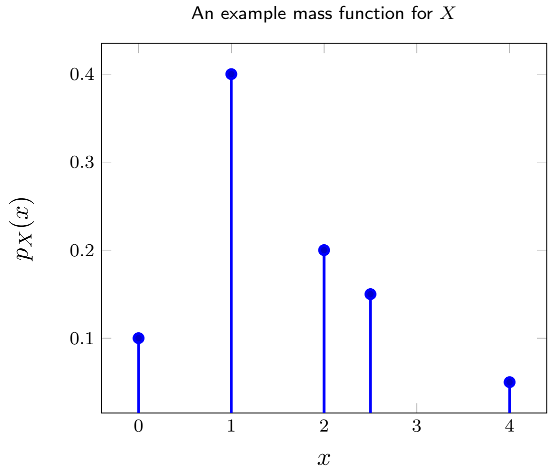
For such discrete random variables, a standard practice is to translate probabilities on outcomes into probabilities on variable values:

$$p_X(x) := P(X = x).$$

p_X is called the *(probability) mass function* for X, and maps elements from the set of values of X to the set of numbers $[0, 1]$.

In particular mass functions have some intuitive properties:

- $p_X(x) = 0$ implies that $X = x$ happens with zero probability.
- The value of $p_X(x)$ is a direct measure of how probable the outcome $X = x$ is.
- $\sum_x p_X(x) = 1$
- $p_X(x) \geqslant 0$



Distribution functions

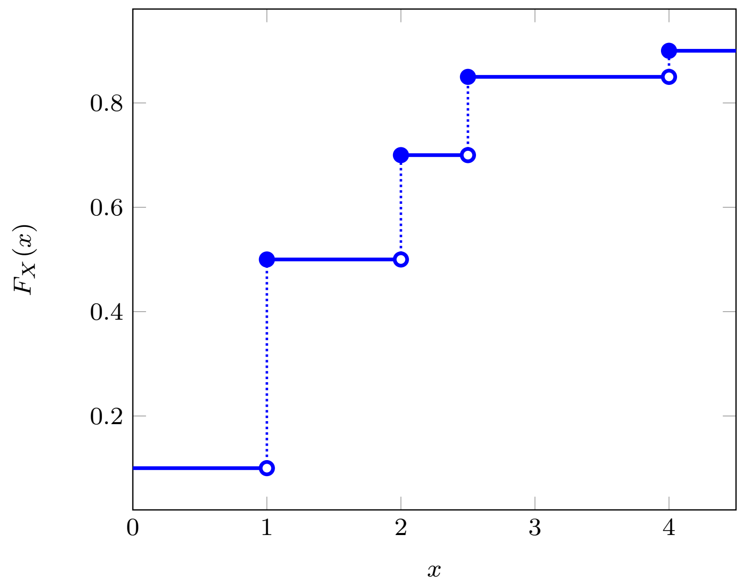
A less obviously useful function is the *(cumulative) distribution function* of X, defined as,

$$F_X(x) := P\left(X \leqslant x\right) = \sum_{y \leqslant x} p_X(y).$$

This measures the (cumulative) probability that X takes on values x or smaller.

This function is monotone non-decreasing, limiting to value 0 as $x \to -\infty$ and to value 1 as $x \to +\infty$.

An example distribution function for X



Statistics, I

With an understanding of the likelihood of outcomes for a random variable X, we can compute *averages*.

The fundamental operator in this sphere is the expectation operator $\mathbb{E}$, acting on a random quantity. This quantity can be an arbitrary well-defined function of a given random variable X:

$$\mathbb{E}g(X) := \sum_x g(x) p_X(x).$$

Intuitively: $p_X(x)$ are convex weights, and hence $\mathbb{E}g(X)$ is a convex combination ("average") of realizations of g.

$w_1 \ldots w_n$ are "convex weights" if $w_i \in [0,1]$, and

$$\sum_{i=1}^{n} w_i = 1$$

Statistics, I

With an understanding of the likelihood of outcomes for a random variable X, we can compute *averages*.

The fundamental operator in this sphere is the expectation operator $\mathbb{E}$, acting on a random quantity. This quantity can be an arbitrary well-defined function of a given random variable X:

$$\mathbb{E}g(X) := \sum_x g(x) p_X(x).$$

Intuitively: $p_X(x)$ are convex weights, and hence $\mathbb{E}g(X)$ is a convex combination ("average") of realizations of g.

We will mostly be concerned with first- and second-order statistics, corresponding to specific choices for g:

$$g(x) = x \quad \longrightarrow \quad \mathbb{E}X = \sum_x x p_X(x) \quad \text{(The \textbf{mean} of } X\text{)}$$

$$g(x) = (x - \mathbb{E}X)^2 \quad \longrightarrow \quad \mathbb{E}(X - \mathbb{E}X)^2 = \sum_x (x - \mathbb{E}X)^2 p_X(x) \quad \text{(The \textbf{variance} of } X\text{)}$$

The mean provides average behavior of X; the variance provides a (coarse) measure of the "spread" of X.

Statistics, II

$\mathbb{E}[X - \mathbb{E}X] = 0$

Some terminology and notation:

- If we choose $g(x) = x^n$, then $\mathbb{E}g(X)$ is typically called the nth (uncentered) moment of X.
- If we choose $g(x) = (x - \mathbb{E}X)^n$, then $\mathbb{E}g(X)$ is typically called the nth centered moment of X.
- The variance of a random variable is often denoted $\mathrm{Var}X := \mathbb{E}(X - \mathbb{E}X)^2 \geqslant 0$.
- The *standard deviation* of X is defined as $\sqrt{\mathrm{Var}X}$.

Statistics, II

Some terminology and notation:

- If we choose $g(x) = x^n$, then $\mathbb{E}g(X)$ is typically called the nth (uncentered) moment of X.
- If we choose $g(x) = (x - \mathbb{E}X)^n$, then $\mathbb{E}g(X)$ is typically called the nth centered moment of X.
- The variance of a random variable is often denoted $\mathrm{Var}X := \mathbb{E}(X - \mathbb{E}X)^2 \geqslant 0$.
- The *standard deviation* of X is defined as $\sqrt{\mathrm{Var}X}$.

There are some properties of these statistics that are reasonably straightforward to show:

- The expectation operator is *linear*: if X and Y are any two random variables, then

$$\mathbb{E}\left(aX + bY\right) = a\mathbb{E}X + b\mathbb{E}Y, \qquad\qquad a, b \in \mathbb{C}$$

- The variance operator is invariant to deterministic shifts, and scales quadratically with scaling:

a deterministic

$$\mathrm{Var}(X + a) = \mathrm{Var}X, \qquad\qquad \mathrm{Var}(aX) = |a|^2\mathrm{Var}X.$$

- The variance of a random variable satisfies:

$$\mathrm{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

Continuous random variables

Most of the story is the same if a random variable X is continuously distributed.

The main difference is that mass functions don't exist/make sense anymore.
E.g., if X is uniformly distributed on $[0, 1]$, then

$$P(X = a) = 0, \qquad\qquad a \in [0, 1],$$

hence the mass function would be zero.

Continuous random variables

Most of the story is the same if a random variable X is continuously distributed.

The main difference is that mass functions don't exist/make sense anymore.
E.g., if X is uniformly distributed on $[0, 1]$, then

$$P(X = a) = 0, \qquad\qquad a \in [0, 1],$$

hence the mass function would be zero.

We "fix" this problem by defining *(probability) density functions*: $f_X(x)$: density

$$P(X \in [a, b]) := \int_a^b f_X(x) \mathrm{d}x, \qquad\qquad a \leqslant b.$$

These types of random variables also have distribution functions:

$$F_X(x) = P(X \leqslant x) = \int_{-\infty}^{x} f_X(y) \mathrm{d}y.$$

Continuous random variables

Most of the story is the same if a random variable X is continuously distributed.

The main difference is that mass functions don't exist/make sense anymore.
E.g., if X is uniformly distributed on $[0, 1]$, then

$$P(X = a) = 0, \qquad\qquad a \in [0, 1],$$

hence the mass function would be zero.

We "fix" this problem by defining *(probability) density functions*:

$$P(X \in [a, b]) := \int_a^b f_X(x) \mathrm{d}x, \qquad\qquad a \leqslant b.$$

These types of random variables also have distribution functions:

$$P(X \leqslant x) = \int_{-\infty}^{x} f_X(x) \mathrm{d}x.$$

Density functions are not quite as transparent as mass functions:

- The value $f_X(x)$ does *not* provide information about the probability that $X = x$.
- While $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$ and $f_X(x) \geqslant 0$, the actual values of $f_X(x)$ can be arbitrarily large numbers.

Continuous random variable statistics

Statistics for continuous random variables is defined essentially the same as for the discrete case.

To see this, we need only define expectation appropriately:

$$\mathbb{E}g(X) = \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x$$

All the definitions and properties of statistics we've seen before are the same.

Conditional probabilities

Conditional probabilities are ways of narrowing the set of events by specifying a condition.

Probabilities must also be renormalized appropriately. Given events A and B, then

$$P(A|B) := \frac{P(A \bigcap B)}{P(B)}.$$

I.e., the probability of A conditioned on B is the probability of *both* A and B happening, normalized by the probability that B happens.

Conditional probabilities

Conditional probabilities are ways of narrowing the set of events by specifying a condition.

Probabilities must also be renormalized appropriately. Given events A and B, then

$$P(A|B) := \frac{P(A \bigcap B)}{P(B)}.$$

I.e., the probability of A conditioned on B is the probability of *both* A and B happening, normalized by the probability that B happens.

Example

I roll a 6-sided fair die.

$$\begin{aligned} P(\text{Face 4 is on top}) &= \frac{1}{6} \\ P(\text{Face 4 is on top} \mid \text{The top face is even}) &= \frac{P(\text{Face 4 is on top and even})}{P(\text{The top face is even})} \\ &= \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}. \end{aligned}$$

Conditional expectations

Conditional probabilities are (actual) probabilities. E.g., consider a discrete RV X and an event A:

$$p_{X|A}(x) := P(X = x \mid A) \in [0, 1]$$

$$\sum_x p_{X|A}(x) = \sum_x P((X = x) \mid A) = \frac{\sum_x P(X = x \bigcap A)}{P(A)} = \frac{P(A)}{P(A)} = 1.$$

Conditional expectations

Conditional probabilities are (actual) probabilities. E.g., consider a discrete RV X and an event A:

$$p_{X|A}(x) := P(X = x \mid A) \in [0,1]$$

$$\sum_x p_{X|A}(x) = \sum_x P((X = x) \mid A) = \frac{\sum_x P(X = x \bigcap A)}{P(A)} = \frac{P(A)}{P(A)} = 1.$$

Hence, one can define a *conditional* expectation operator:

$$\mathbb{E}[g(X) \mid A] = \sum_x g(X) p_{X|A}(x).$$

With this, one can define conditional means, variances, etc.

Random vectors

Everything we've discussed essentially generalizes appropriately to vector-valued random variables:

$$\boldsymbol{X} = \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix}^T \in \mathbb{R}^n$$

E.g., if X is discrete, then its mass function $p_{\boldsymbol{X}}$ is a function defined on n-dimensional vectors:

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x}) = P\left((X_1 = x_1)\bigcap(X_2 = x_2)\bigcap\cdots\bigcap(X_n = x_n)\right).$$

Random vectors

Everything we've discussed essentially generalizes appropriately to vector-valued random variables:

$$\boldsymbol{X} = \begin{pmatrix} X_1 & X_2 & \cdots & X_n \end{pmatrix}^T \in \mathbb{R}^n$$

E.g., if X is discrete, then its mass function $p_{\boldsymbol{X}}$ is a function defined on n-dimensional vectors:

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x}) = P\left((X_1 = x_1) \bigcap (X_2 = x_2) \bigcap \cdots \bigcap (X_n = x_n)\right).$$

Hence, the expectation operator is defined in exactly the same manner:

$$\mathbb{E} g(\boldsymbol{X}) = \sum_{\boldsymbol{x}} g(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}),$$

so that the (vector-valued) mean is well-defined:

$$\mathbb{E}\boldsymbol{X} = \sum_{\boldsymbol{x}} \boldsymbol{x} p_{\boldsymbol{X}}(\boldsymbol{x}) \in \mathbb{R}^n.$$

Covariances

There is a hiccup when it comes to second-order statistics of random vectors: which quadratic function should we take the expectation of?

Covariances

There is a hiccup when it comes to second-order statistics of random vectors: which quadratic function should we take the expectation of?

"All of them" is the somewhat unsatsifying answer.

First, the moment of the product of centered versions of X_i and X_j is the *covariance* of X_i and X_j:

$$\mathrm{Cov}(X_i, X_j) := \mathbb{E}\left[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\right].$$

Second, if $\boldsymbol{X} \in \mathbb{R}^n$, then

$$\boldsymbol{X}\boldsymbol{X}^T = \begin{pmatrix} X_1 X_1 & X_1 X_2 & \cdots & X_1 X_n \\ X_2 X_1 & X_2 X_2 & \cdots & X_2 X_n \\ \vdots & & \ddots & \vdots \\ X_n X_1 & \cdots & & X_n X_n \end{pmatrix} \in \mathbb{R}^{n \times n},$$

is a matrix containing every quadratic combination of the components of $\boldsymbol{X}$.

Covariances

There is a hiccup when it comes to second-order statistics of random vectors: which quadratic function should we take the expectation of?

"All of them" is the somewhat unsatsifying answer.

First, the moment of the product of centered versions of X_i and X_j is the *covariance* of X_i and X_j:

$$\mathrm{Cov}(X_i, X_j) := \mathbb{E}\left[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\right].$$

Second, if $\boldsymbol{X} \in \mathbb{R}^n$, then

$$\boldsymbol{X}\boldsymbol{X}^T = \begin{pmatrix} X_1X_1 & X_1X_2 & \cdots & X_1X_n \\ X_2X_1 & X_2X_2 & \cdots & X_2X_n \\ \vdots & & \ddots & \vdots \\ X_nX_1 & \cdots & & X_nX_n \end{pmatrix} \in \mathbb{R}^{n\times n},$$

is a matrix containing every quadratic combination of the components of $\boldsymbol{X}$.

With this setup, the *covariance matrix* of $\boldsymbol{X}$ the matrix of covariances between components of $\boldsymbol{X}$:

$$\mathrm{Cov}(\boldsymbol{X}) := \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})(\boldsymbol{X} - E\boldsymbol{X})^T\right].$$

The covariance matrix

$$\mathrm{Cov}(\boldsymbol{X}) := \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})(\boldsymbol{X} - E\boldsymbol{X})^T\right].$$

Some direct consequences:

1. $\mathrm{Cov}(\boldsymbol{X})$ is symmetric
2. $\mathrm{Cov}(\boldsymbol{X})$ is positive ~~definite~~ semidefinite
3. For a deterministic $\boldsymbol{a} \in \mathbb{R}^n$, $\mathrm{Var}\langle \boldsymbol{a}, \boldsymbol{X}\rangle = \boldsymbol{a}^T \mathrm{Cov}(\boldsymbol{X})\boldsymbol{a}$.

$A = \mathrm{Cov}(X)$

$w \in \mathbb{R}^n$ $\quad \mu = \mathbb{E}X$

$w^T A w = w^T \mathbb{E}[(X-\mu)(X-\mu)^T] w$

$= \mathbb{E}[w^T(X-\mu)(X-\mu)^T w]$

$= \mathbb{E}[(w^T(X-\mu))^2] \geq 0$

The covariance matrix

$$\text{Cov}(\boldsymbol{X}) := \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})(\boldsymbol{X} - E\boldsymbol{X})^T\right].$$

Some direct consequences:

1. $\text{Cov}(\boldsymbol{X})$ is symmetric
2. $\text{Cov}(\boldsymbol{X})$ is positive definite
3. For a deterministic $\boldsymbol{a} \in \mathbb{R}^n$, $\text{Var}\left\langle \boldsymbol{a}, \boldsymbol{X} \right\rangle = \boldsymbol{a}^T \text{Cov}(\boldsymbol{X}) \boldsymbol{a}$.

Other properties:

The diagonal element $(\text{Cov}(\boldsymbol{X}))_{j,j}$ equals $\text{Var} X_j$.

The *scaled* off-diagonal entries are called (Pearson) *correlation coefficients*:

$$\text{Corr}(X_i, X_j) := \frac{(\text{Cov}(\boldsymbol{X}))_{j,j} \; i,j}{\sqrt{(\text{Var}X_i)(\text{Var}X_j)}} \in [-1, 1]$$

Values "close" to $+1$ indicate that X_i and X_j are "correlated".
Values "close" to -1 indicate that X_i and X_j are "anti-correlated".
A value of 0 indicates that X_i and X_j are "uncorrelated".

The covariance matrix

$$\mathrm{Cov}(\boldsymbol{X}) := \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})(\boldsymbol{X} - E\boldsymbol{X})^T\right].$$

Some direct consequences:

1. $\mathrm{Cov}(\boldsymbol{X})$ is symmetric
2. $\mathrm{Cov}(\boldsymbol{X})$ is positive definite
3. For a deterministic $\boldsymbol{a} \in \mathbb{R}^n$, $\mathrm{Var}\left\langle \boldsymbol{a}, \boldsymbol{X} \right\rangle = \boldsymbol{a}^T \mathrm{Cov}(\boldsymbol{X})\boldsymbol{a}$.

Other properties:

The diagonal element $(\mathrm{Cov}(\boldsymbol{X}))_{j,j}$ equals $\mathrm{Var} X_j$.

The *scaled* off-diagonal entries are called (Pearson) *correlation coefficients*:

$$\mathrm{Corr}(X_i, X_j) := \frac{(\mathrm{Cov}(\boldsymbol{X}))_{j,j}}{\sqrt{(\mathrm{Var} X_i)(\mathrm{Var} X_j)}} \in [-1, 1]$$

Values "close" to $+1$ indicate that X_i and X_j are "correlated".
Values "close" to -1 indicate that X_i and X_j are "anti-correlated".
A value of 0 indicates that X_i and X_j are "uncorrelated".

Uncorrelated random variables are generally not independent: independence requires

$$P((X \in S) \bigcap (Y \in T)) = P(X \in S) P(Y \in T)$$

for all sets S, T. If X and Y are independent they must be uncorrelated, but the reverse need not be true.

Parametric distributions

Some examples of "important" probability distributions are:

- Discrete random variables
  - Bernoulli
  - discrete uniform
  - Binomial
  - Poisson
  - $\vdots$
- Continuous random variables
  - Uniform
  - Beta
  - Gaussian
  - Exponential
  - $\vdots$

We'll discuss various canonical probability distributions throughout this course.

Questions + comments

Suppose X and Y share the same mean and variance. Does $X = Y$?

No:
$$X = \begin{cases} 1, & \text{Prob } 1/2 \\ -1, & \text{Prob } 1/2 \end{cases}$$

$$Y = -X$$

Questions + comments

Suppose X and Y share the same mean and variance, and $\mathrm{Corr}(X,Y) = 1$. Does $X = Y$?

Yes

$\Downarrow$

Strict

linear relationship.

Questions + comments

Suppose X and Y are discrete RV's with the same mass function, i.e., $p_X(m) = p_Y(m)$ for all m. Does $X = Y$?

No

Questions + comments

Let X be a random variable. Does X have either a mass function or a density function?

No (only has a distribution)

Ex: Sum a discrete and a continuous RV.

Questions + comments

Suppose $\mathbb{E}X = \mu$ and $\mathrm{Var}(X) = \sigma^2$. Then is $X \in [\mu - \sigma, \mu + \sigma]$ say with some predictable probability?

Not really.

But, "most" of the time this is "ok".

Questions + comments

Suppose we pick two stocks with the same price today. Tomorrow, we model these share prices as random variables X and Y, with $\mathbb{E}X = \mathbb{E}Y$ and $\mathrm{Var}X < \mathrm{Var}Y$.

Would you advise an investor to invest in stock X instead of stock Y?

References I