MIT MANAGEMENT
SLOAN SCHOOL

15.455x Mathematical Methods of Quantitative Finance

Week 1: Probability

Paul F. Mende
MIT Sloan School of Management

Finance at MIT
Where ingenuity drives results

Random variables, distributions, and moments

Finance at MIT
Where ingenuity drives results

Random variables

future outcomes or observations

- A **random variable** is a function which assigns a number to **events** in the **sample space**. (A better name might be "random-valued function on the sample space.")

sample space is a set of all events, or possible outcomes of experiments, or possible observations that we might have

$$X = \begin{cases} 1, & \text{heads} \\ -1, & \text{tails} \end{cases} \qquad Z = \begin{cases} 1, & \text{success} \\ 0, & \text{failure} \end{cases}$$

$$Y = \{\text{sum of two dice}\}$$

- We describe the **probability of an outcome** in terms of the probability of a random variable taking a given value:

$$P(X=1) = 1/2, \quad P(X^2=1) = 1,$$
$$P(Y=2) = 1/36, \quad P(Y=3) = 2/36, \ldots$$
$$P(Z=1) = p, \quad P(Z=0) = q, \quad (\text{where } p+q=1)$$

Continuous random variables

- Consider choosing a random number between 0 and 1, where all values are equally likely.

- Since there are an (uncountably) **infinite** number of values, the probability of any given value is zero:

$$\mathrm{Prob}(X = x_0) = 0$$

- Does **not** mean event is impossible. Points on a line have **"measure zero."**

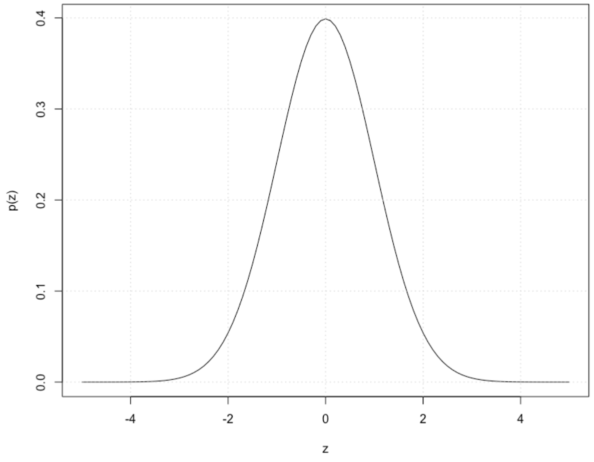
- Instead, ask for probability to lie within a given range, *e.g.*,

$$\mathrm{Prob}(a < X < b) = \int_a^b p(x)\mathrm{d}x$$

Probability distributions

- More generally, a **probability distribution** satisfies

$$
\begin{aligned}
p(x) &\geq 0, \\
\sum_k p(x_k) &= 1, \quad x_k \text{ discrete} \\
\int_{-\infty}^{\infty} p(x)\mathrm{d}x &= 1, \quad x \text{ continuous}
\end{aligned}
$$



Cumulative distribution function

- The **cumulative distribution function**, or CDF, gives the probability that *X* is less than or equal to a given value.

$$F(x) = \mathrm{Prob}(X < x) = \int_{-\infty}^{x} p(x')\mathrm{d}x'$$

- It contains (nearly) the same information as the probability density, since

$$p(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$$

- The CDF is often easier to approximate from empirical data and it is useful since

$$\mathrm{Prob}(a < X < b) = F(b) - F(a)$$

Change of variable

- Suppose we wish to change variables, shift the distribution, or consider functions of the random variable. If $x \to y = y(x)$ then the density in terms of the new variable is given by

$$p(x)\mathrm{d}x = g(y)\mathrm{d}y$$

which preserves the normalization condition. (This assumes *y* is an increasing function of *x.* If not, an absolute value is needed for positivity; and further care is needed if *y* has critical points.)

$$\begin{aligned} p(x)\mathrm{d}x &= g(y)\mathrm{d}y \\ g(y) &= \frac{p(x)}{|\mathrm{d}y/\mathrm{d}x|} \end{aligned}$$

Expectations and moments

- The probability distribution defines weighted averages over the sample space, where the weight of each event is equal to its probability. These are called **expected values**.

- For the discrete case,

$$\mathrm{E}\left[f(X)\right] = \sum_{k=1}^{n} f(x_k)p(x_k)$$

while for the continuous case,

$$\mathrm{E}\left[f\right] = \int_{-\infty}^{\infty} f(x)p(x)\mathrm{d}x$$

Mean of a distribution

- The **mean** of the distribution is simply the expectation of the random variable itself:

$$\mu \equiv \mathrm{E}\left[X\right] = \overline{X} = \langle X \rangle = \begin{cases} \sum_k x_k p(x_k) \\ \int x p(x) \mathrm{d}x \end{cases}$$

- In the case of an infinite sample space, whether continuous or discrete, the mean is not guaranteed to exist since the integral or the sum might not converge.

Moments of a distribution

- The **moments** of a distribution are the expectation of powers of the random variable itself.

$$\mu_\ell \equiv \mathrm{E}\left[X^\ell\right] \equiv \langle X^\ell \rangle = \begin{cases} \sum_k x_k^\ell p(x_k) \\ \int x^\ell p(x)\mathrm{d}x \end{cases}$$

might not exist in the case of a continuous distribution.

- If all the moments are known – and if they exist – they can be used to get the expectation of other functions using the **linearity** of the expectation operator

$$\mathrm{E}\left[cf(X)\right] = c\,\mathrm{E}\left[f(X)\right],$$
$$\mathrm{E}\left[f(X) + g(X)\right] = \mathrm{E}\left[f(X)\right] + \mathrm{E}\left[g(X)\right]$$

if we know all of the moments, then we can reconstruct the full probability distribution itself. (Taylor's theorem)

Variance and standard deviation

- Of particular interest is the second moment, in combination with the mean, defining the **variance**:

$$
\begin{aligned}
\sigma^2 = \mathrm{Var}(X) \equiv \mathrm{E}\left[(X-\mu)^2\right] &= \mathrm{E}\left[X^2 - 2\mu X + \mu^2\right] \\
&= \mathrm{E}\left[X^2\right] - 2\mu \mathrm{E}\left[X\right] + \mu^2 \\
&= \mathrm{E}\left[X^2\right] - \mu^2 \\
&= \mathrm{E}\left[X^2\right] - \mathrm{E}\left[X\right]^2
\end{aligned}
$$

volatility in financial processes

- The standard deviation, which is the square root of the variance, has the **same units** as the random variable (e.g., rate of return, dollars, etc.)

Higher moments characterize properties of a distribution

Variance – dispersion measure based on second moment

$$\sigma^2 \equiv \mathrm{E}\left[(X-\mu)^2\right] = \int (x-\mu)^2\, p(x)\mathrm{d}x$$

Skewness – asymmetry parameter based on 3rd moments; **dimensionless** – normalized cumulant a skewness of 0 is what we'll find if a probability distribution is symmetric.

$$s \equiv \frac{\mathrm{E}\left[(X-\mu)^3\right]}{\sigma^3} = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

Kurtosis – measure of tail "weights" in terms of 4th moments; zero for Gaussian, bounded below by -1.
normal distribution

$$\kappa \equiv \boxed{\frac{\mathrm{E}\left[(X-\mu)^4\right]}{\sigma^4}} - 3 \qquad \text{excess kurtosis}$$

non–excess kurtosis

Covariance and correlation

MIT MANAGEMENT SLOAN SCHOOL

- For any **two** random variables, not necessarily independent **or** identically distributed, their covariance is defined as

not dimensionless
$$\mathrm{Cov}(X,Y) \equiv \mathrm{E}\left[(X-\mu_x)(Y-\mu_y)\right] = \mathrm{E}\left[XY\right] - \mu_x\mu_y$$

- The correlation is proportional to the covariance,

$$\rho(X,Y) = \mathrm{Corr}(X,Y) \equiv \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \mathrm{E}\left[\left(\frac{X-\mu_x}{\sigma_x}\right)\left(\frac{Y-\mu_y}{\sigma_y}\right)\right]$$

dimensionless

- Dividing the covariance by the standard deviations makes the correlation a pure number, and

$$-1 \leq \rho(X,Y) \leq +1$$

Covariance

- If the variables are independent, then the covariance will vanish. However the converse is not true.
  - Example: Let Y be a function of X, so it is completely <u>dependent</u>:

$$X = \pm a \text{ each with probability } p_1,\ \pm b \text{ with probability } p_2$$
$$Y = X^2 = a^2 \text{ with probability } 2p_1,\ b^2 \text{ with probability } 2p_2$$
$$\mu_x = 0, \quad \mu_y = \mathrm{E}[Y] = 2p_1 a^2 + 2p_2 b^2$$

$$\begin{aligned}\mathrm{Cov}(X,Y) &\equiv \mathrm{E}[(X-\mu_x)(Y-\mu_y)] = \mathrm{E}[XY] - \mu_x\mu_y = \mathrm{E}[XY] \\ &= p_1 \cdot a \cdot a^2 + p_1(-a)\cdot a^2 + p_2 \cdot b \cdot b^2 + p_2(-b)\cdot b^2 \\ &= 0\end{aligned}$$

  - Vanishing covariance does not imply independence.

Random variables

Summary

a number, not probability

- "**Random variables**" are functions that assign **a number** to events in the sample space. They can be discrete, continuous, or a mix of both.

- The probability distribution is positive and sums to one.

- Expected values, or expectations, are weighted averages that use the probabilities as the weights.

- The **moments** of a distribution, such as the mean, are expectations of various powers of a random variable. They are numbers, not functions.

- Variance, skewness, and kurtosis are simple functions of the moments that characterize the shape of the probability distribution.

- When there are multiple random variables, their **covariance** and **correlation** are also computed as expectations.

Common distributions

Finance at MIT
Where ingenuity drives results

Uniform distribution

- Standard form: $$p(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & \text{otherwise} \end{cases},$$

$$\mathrm{Prob}(a < X < b) = \int_a^b p(x)\mathrm{d}x = (b-a)$$

$$F(x) = \int_{-\infty}^{x} p(x')\mathrm{d}x' = x$$

- Moments:

$$\mu = \int_{-\infty}^{\infty} x\, p(x)\mathrm{d}x = \int_0^1 x\mathrm{d}x = \frac{1}{2}$$

$$\mu_\ell = \int_0^1 x^\ell \mathrm{d}x = \frac{1}{\ell+1}$$

$$\sigma^2 = \int_0^1 \left(x - \frac{1}{2}\right)^2 \mathrm{d}x = \frac{1}{12}$$

Binomial distribution

- The **binomial distribution** provides a model for any observation or experiment that has two possible outcomes, "success" or "failure."

- Examples:
  - Coin toss: Heads vs. Tails
  - Roulette: 19 vs. {all other numbers}
  - Bond default prior to time *T* vs. no default
  - One-day trading loss > $100M vs. (P/L > -$100M)

Binomial distribution

- Identify events by total number of successes vs. failures, independent of the order in which they occur.

- **Two parameters**: probability of success *p*, number of trials *n.*

$$f(k; n, p) = \binom{n}{k} p^k q^{n-k}, \text{ where}$$

$$q = 1 - p \text{ is probability of ``failure,''}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$$

Binomial distribution: moments

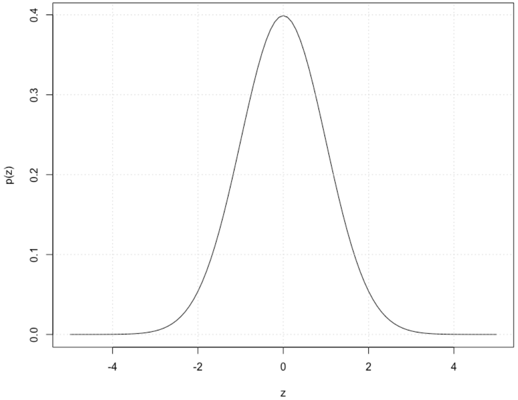
- Mean value (the hard way):

$$
\begin{aligned}
\mu = \mathrm{E}\left[X\right] &= \sum_{k=0}^{\infty} k\, f(k; n, p) \\
&= \sum_{k=1}^{n} \frac{k\, n!}{k!(n-k)!} p^k q^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{(n-1)-(k-1)} \\
&= np \sum_{k'=0}^{n'} f(k'; n', p), \;\; \text{where } k' = k-1, \;\; n' = n-1 \\
&= np
\end{aligned}
$$

Gaussian distribution

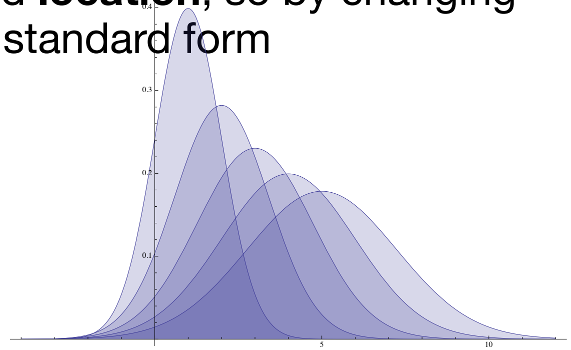
- The **normal** distribution – also known as **Gaussian** – is defined by the two-parameter probability density function

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$



- Actually, the two parameters give its **scale** and **location**, so by changing to a rescaled variable, all Gaussians have the standard form

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \text{ where } z = \frac{x-\mu}{\sigma}$$



Gaussian distribution

- Check the normalization, mean, and variance with handy formulas

$$I(a) = \int_{-\infty}^{\infty} e^{-ax^2} \mathrm{d}x = \sqrt{\frac{\pi}{a}}$$

$$-\frac{\mathrm{d}I(a)}{\mathrm{d}a} = \int_{-\infty}^{\infty} x^2 e^{-ax^2} \mathrm{d}x = \frac{1}{2}\sqrt{\frac{\pi}{a^3}}$$

move the derivative inside the integral

- Probability in a range computed from CDF,

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z'^2/2} \mathrm{d}z' = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$F(b) - F(a) = \mathrm{Prob}(a < Z < b)$$

Lognormal distribution

- If we change variables and substitute, obtain the **lognormal** density function

$$X = \log Y, \quad x = \log y,$$
$$\text{then } g(y) = \frac{p(x)}{|\mathrm{d}y/\mathrm{d}x|} = \frac{1}{y\sqrt{2\pi\sigma^2}} e^{-(\log y - \mu)^2/(2\sigma^2)}$$

- The moments are more easily computed in original variables,

R: simple returns on an asset

$$\mathrm{E}[R] = \mathrm{E}[e^r - 1] = e^{\mu+\sigma^2/2} - 1 \approx \mu + \sigma^2/2 + \ldots,$$
$$\mathrm{Var}(R) = \mathrm{E}\left[(R-\bar{R})^2\right] = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) \approx \sigma^2(1+2\mu) + \ldots$$

Poisson distribution

- Depends on two parameters, one integer and one real:

$$p(k; \lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$$

arrival rate

- Application: probability to find exactly *k* "arrivals" or "events" during interval of time *t*.

$$p(k; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

Poisson distribution

- Mean

$$\mu = \mathrm{E}\left[X\right] = e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda\, e^{-\lambda} \sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!} = \lambda$$

- Second moment

$$\mathrm{E}\left[X^2\right] = e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} = \lambda + \lambda^2,$$

$$\sigma^2 = \mathrm{E}\left[X^2\right] - \mu^2 = \lambda$$

Poisson distribution

- Limiting case of binomial distribution $n \to \infty, \quad np = \lambda$ fixed

$$p(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$$

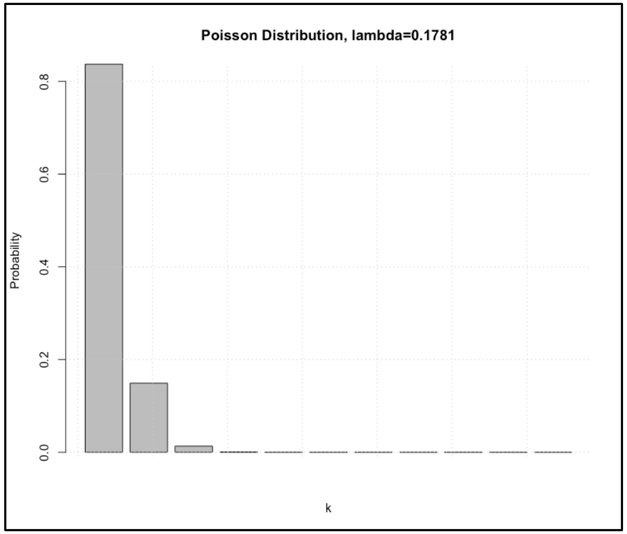
- Example: how many students in a room have a birthday today?

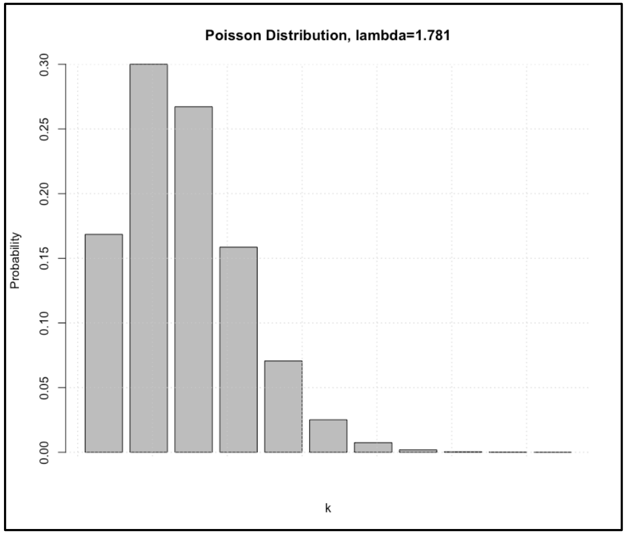
$$p = \frac{1}{365}, \; n = 65$$
$$\lambda = np \approx 0.1781$$

Number *k*	Binomial probability	Poisson probability
0	0.8367	0.8369
1	0.1494	0.1490
2	0.0131	0.0133

Poisson distribution

- For small lambda, values are always decreasing in *k*.
- For lambda > 1, there is a single maximum for *k* near lambda
- Distribution is **skewed** with a long right-hand tail.





Poisson distribution

- A few more handy formulas for evaluating moments:

$$
\begin{aligned}
\frac{k}{k!} &= \frac{k}{k(k-1)!} = \frac{1}{(k-1)!}, \\
\frac{k^2}{k!} &= \frac{k}{(k-1)!} = \frac{(k-1)+1}{(k-1)!} = \frac{1}{(k-2)!} + \frac{1}{(k-1)!}, \\
\frac{k^3}{k!} &= \frac{k^2}{(k-1)!} = \frac{(k^2-3k+2)+(3k-3)+1}{(k-1)!} = \frac{1}{(k-3)!} + \frac{3}{(k-2)!} + \frac{1}{(k-1)!}
\end{aligned}
$$

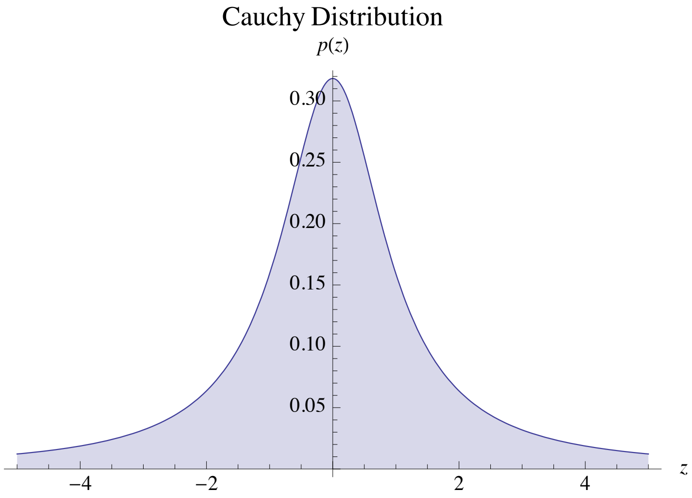
- To simplify these ratios, add and subtract terms terms in the numerators that help simplify the denominators, always starting from the highest power of *k* and working down

- Is there a more general way to generate **all the moments** of a distribution?

Fat tails

- The term "fat tails" is sometimes used to refer to distributions with a large kurtosis.
- It may also refer to distributions for which **higher moments diverge**, in which case the CLT may not hold.
- Example: power-law tails
  - Cauchy distribution

$$p(x) = \frac{A}{\pi^2 A^2 + x^2} \sim \frac{A}{x^2}$$



  - Variance and higher moments are infinite

none of the moments exist

Common distributions

Summary

- Common probability distributions used in finance include
  - Uniform
  - Binomial
  - Poisson
  - Normal (Gaussian)
  - Lognormal
- These have well-defined moments and interesting limiting cases
- Applications include option pricing, credit defaults, jump processes, and most models of asset pricing and forecasting.
- Choosing the right distribution depends on theory (sometimes) and on empirical observations (always).

Sums of random variables

Finance at MIT
Where ingenuity drives results

Multiple random variables

- When there are several random variables, it is often useful to ask about the **distribution of their sum.**

$$S = X_1 + X_2 + \cdots + X_n$$

- In simple cases, the full probability distribution can be derived.

- In more complex cases, we only try to compute some of the **moments** of the full distribution, such as the mean and the variance.
- Easy to do using **linearity** of the expectation operator

$$\mathrm{E}[S] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n]$$

- Due to the CLT, this is often enough! If there are enough variables, and if the individual distributions in the sum are well-behaved, then the sum will be an (approximately) Gaussian variable…in which case the mean and variance describe it completely. regardless of what the X initial distributions are

Application: portfolio returns

- Financial modeling applications include sums over assets and sums across time. Their **risk** depends on the **covariance**.
  - Example: Let a portfolio consist of N assets and let
    - w_i = portfolio weight in asset i,
    - R_i be the return on asset i,
    - $\mu_i = \mathrm{E}[R_i], \quad \sigma_i^2 = \mathrm{Var}(R_i)$

$$R_p = \sum_{i=1}^{N} w_i R_i = w_1 R_1 + w_2 R_2 + \cdots + w_N R_N$$

$$\mu_p = \mathrm{E}[R_p] = \sum_{i=1}^{N} w_i \mathrm{E}[R_i] = \sum_{i=1}^{N} w_i \mu_i$$

Correlated random variables

- No assumptions made about the individual return distributions except for the existence of the mean and variance.

- The portfolio risk, measured by its standard deviation, does involve the **covariance of pairs** of asset returns

$$
\begin{aligned}
\sigma_p^2 = \mathrm{Var}(R_p) &= \mathrm{E}\left[(R_p - \mu_p)^2\right] = \mathrm{E}\left[\left(\sum_{i=1}^{N} w_i (R_i - \mu_i)\right)^2\right] \\
&= \sum_{i=1}^{N} w_i^2\, \mathrm{E}\left[(R_i - \mu_i)^2\right] + 2\sum_{i<j}^{N} w_i w_j\, \mathrm{E}\left[(R_i - \mu_i)(R_j - \mu_j)\right] \\
&= \sum_{i=1}^{N} w_i^2 \mathrm{Var}(R_i) + 2\sum_{i<j}^{N} w_i w_j \mathrm{Cov}(R_i, R_j) = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2\sum_{i<j}^{N} w_i w_j \sigma_i \sigma_j \rho_{ij}
\end{aligned}
$$

Correlated random variables: special cases

- Correlation = 0, variance equals sum of squares

$$\rho_{ij} = 0, \qquad \sigma_p^2 = \sum (w_i \sigma_i)^2$$

- Correlation = 1, variance equals square of sum

$$\rho_{ij} = 1, \qquad \sigma_p^2 = \left(\sum w_i \sigma_i\right)^2$$

- Correlation = 0, all weights and variances identical

$$\rho_{ij} = 0, w_i = \frac{1}{N}, \sigma_i = \sigma_0 \quad \Longrightarrow \quad \sigma_p^2 = \frac{\sigma_0^2}{N}$$

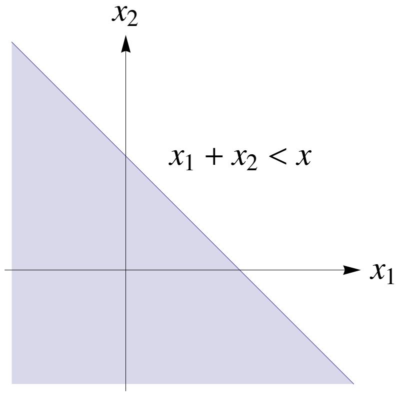
That means that as n gets large, the volatility in the portfolio goes down. And this is the basic idea behind portfolio diversification.

Two random variables

- Two, however, is not a large number.

Let $S = X_1 + X_2$ be the sum of two independent random variables with density function $p_1(x_1)$ and $p_2(x_2)$. Then

$$F(x) = \mathrm{Prob}(X_1 + X_2 < x) = \int\int_A p_1(x_1)p_2(x_2)\mathrm{d}x_1\mathrm{d}x_2$$
$$= \int_{x_1=-\infty}^{x_1=\infty}\int_{x_2=-\infty}^{x_2=x-x_1} p_1(x_1)p_2(x_2)\mathrm{d}x_1\mathrm{d}x_2$$

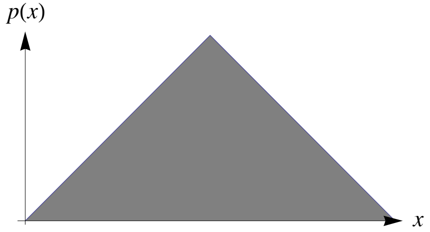


So

$$p(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x} = \int_{-\infty}^{\infty} p_1(x_1)p_2(x - x_1)\mathrm{d}x_1,$$

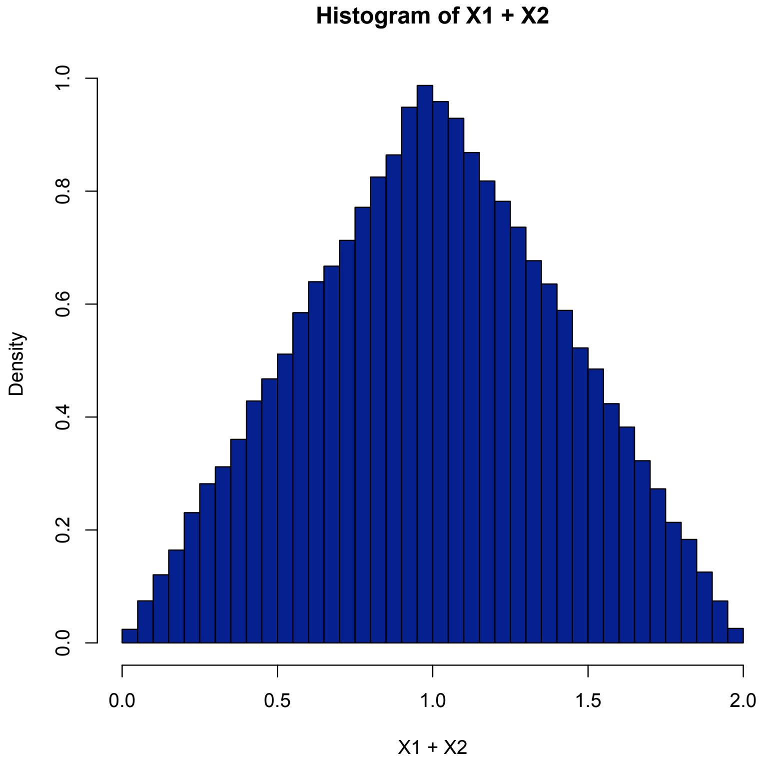
which defines the *convolution* of p_1 and p_2. In the case of the uniform distribution

$$p(x) = \begin{cases} x & x \in [0,1] \\ 2 - x & x \in [1,2] \\ 0 & \text{otherwise} \end{cases}$$



Two random variables: simulation

```
X1 <- runif(1e5)    # X1 and X2 drawn from uniform distribution on [0,1]
X2 <- runif(1e5)
hist(X1+X2, probability=TRUE, breaks=50)
```



Binomial distribution (the easy way)

- We can view the number of "successes" as a **sum of random variables**, one for each of the Bernoulli trials,

$$
\begin{aligned}
\text{Let } S &= X_1 + X_2 + \ldots + X_n. \\
\text{Then } \mathrm{E}[S] &= \mathrm{E}[X_1] + \mathrm{E}[X_2] + \ldots + \mathrm{E}[X_n] \\
&= np.
\end{aligned}
$$

- This approach builds up results, using the expectations of one or two of the random variables at a time and the fact that the *X*'s are **independent** and **identically distributed.**

$$
\begin{aligned}
\mathrm{E}[X_1] &= p \cdot 1 + q \cdot 0 = p, \\
\mathrm{E}\left[X_1^2\right] &= p \cdot 1^2 + q \cdot 0^2 = p, \\
\mathrm{E}[X_1 X_2] &= p^2 \cdot 1 \cdot 1 + pq \cdot 1 \cdot 0 + qp \cdot 0 \cdot 1 + q^2 \cdot 0 \cdot 0 = p^2
\end{aligned}
$$

Binomial distribution

- This approach makes the variance especially easy to compute:

$$
\begin{aligned}
\sigma^2 &= \mathrm{E}\left[(S-\mu)^2\right] \\
&= \mathrm{E}\left[(X_1 + X_2 + \ldots + X_n)^2\right] - \mu^2 \\
&= n\,\mathrm{E}\left[X_1^2\right] + n(n-1)\,\mathrm{E}\left[X_1 X_2\right] - \mu^2 \\
&= np + n(n-1)p^2 - (np)^2, \;\; \text{since } \mathrm{E}\left[X_i X_j\right] = \begin{cases} p, & \text{if } i = j \\ p^2, & \text{if } i \neq j \end{cases} \\
&= np(1-p) = npq,
\end{aligned}
$$

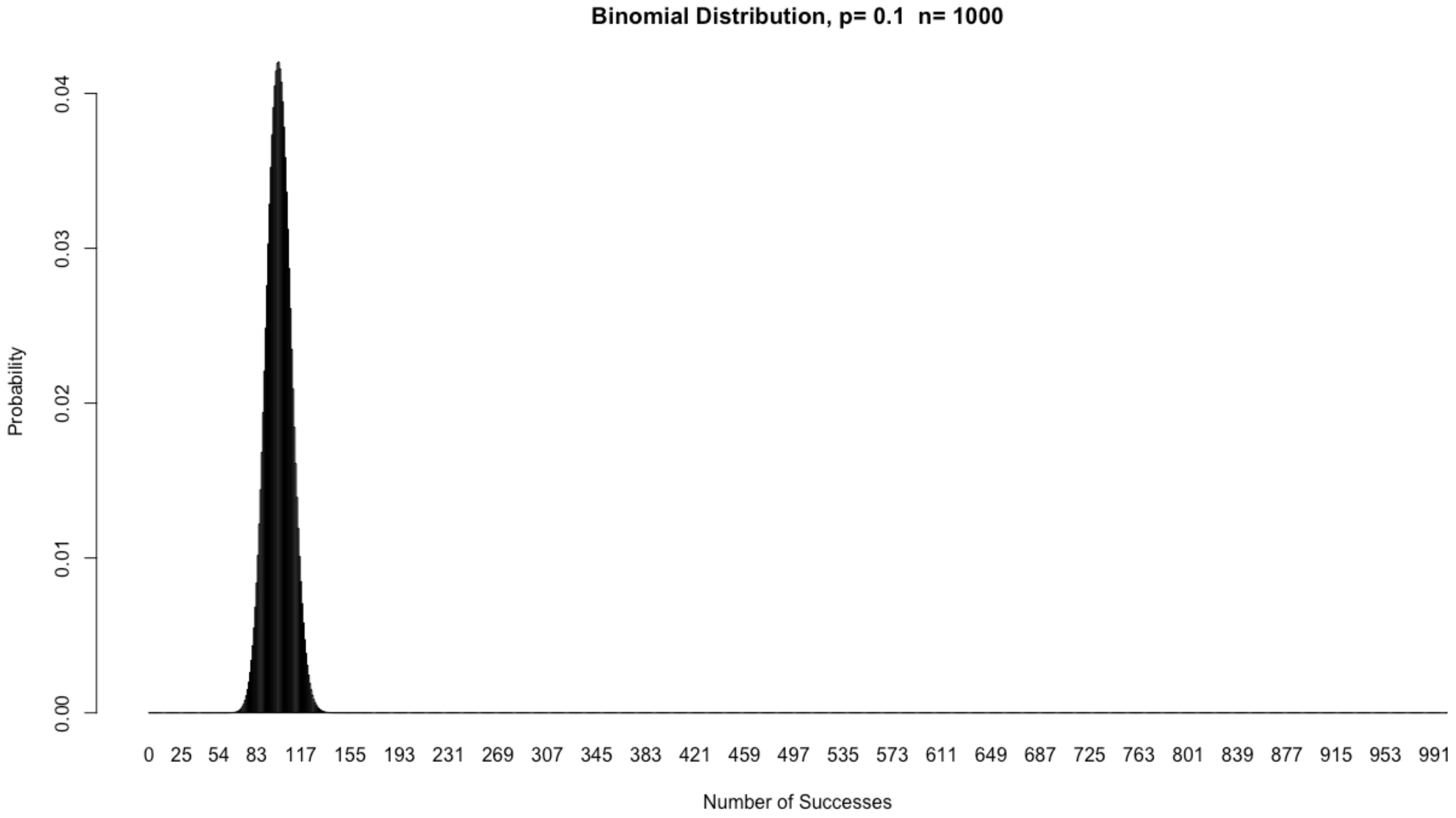
$$
\sigma = \sqrt{npq}.
$$

Binomial distribution

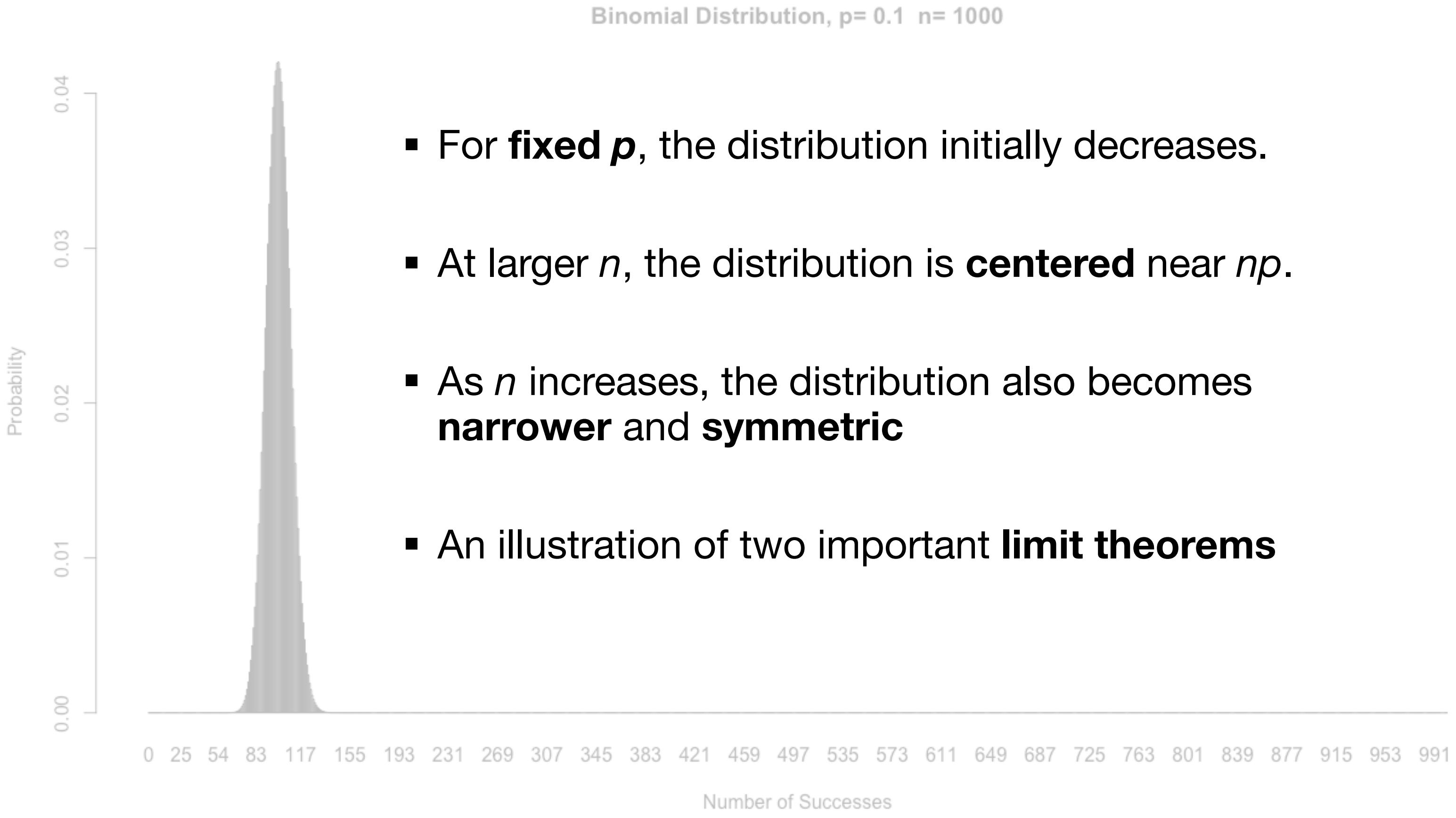
- In numerical calculations, factorial factors quickly overflow machine limits
- R functions:
  - choose(n,k)
  - dbinom(k,n,p)

```
nlist <- c(1,2,5,10,20,50,100,1000)
p <- 0.1
for (n in nlist) {
          k <- 0:n
          f <- dbinom(k,n,p)
          barplot(f,names=k,
          xlab="Number of Successes",ylab="Probability",
          main=paste("Binomial Distribution, p=",p," n=",n))
          readline()
}
```

Binomial distribution, p=10%, n=1,2,...,1000



Binomial distribution



- For **fixed *p***, the distribution initially decreases.
- At larger *n*, the distribution is **centered** near *np*.
- As *n* increases, the distribution also becomes **narrower** and **symmetric**
- An illustration of two important **limit theorems**

Binomial distribution

An illustration of two important **limit theorems**

- **Law of Large Numbers (LLN):**
  - As n increases, the probability that the mean deviates from np goes to zero.

that's why the probability distribution is narrowing around that mean value.

- **Central Limit Theorem (CLT):**
  - As n increases for fixed p, the distribution approaches Gaussian (or Normal)

a bunch of non– Gaussian random variables, when we add 1,000 of them together, we get something that looks Gaussian

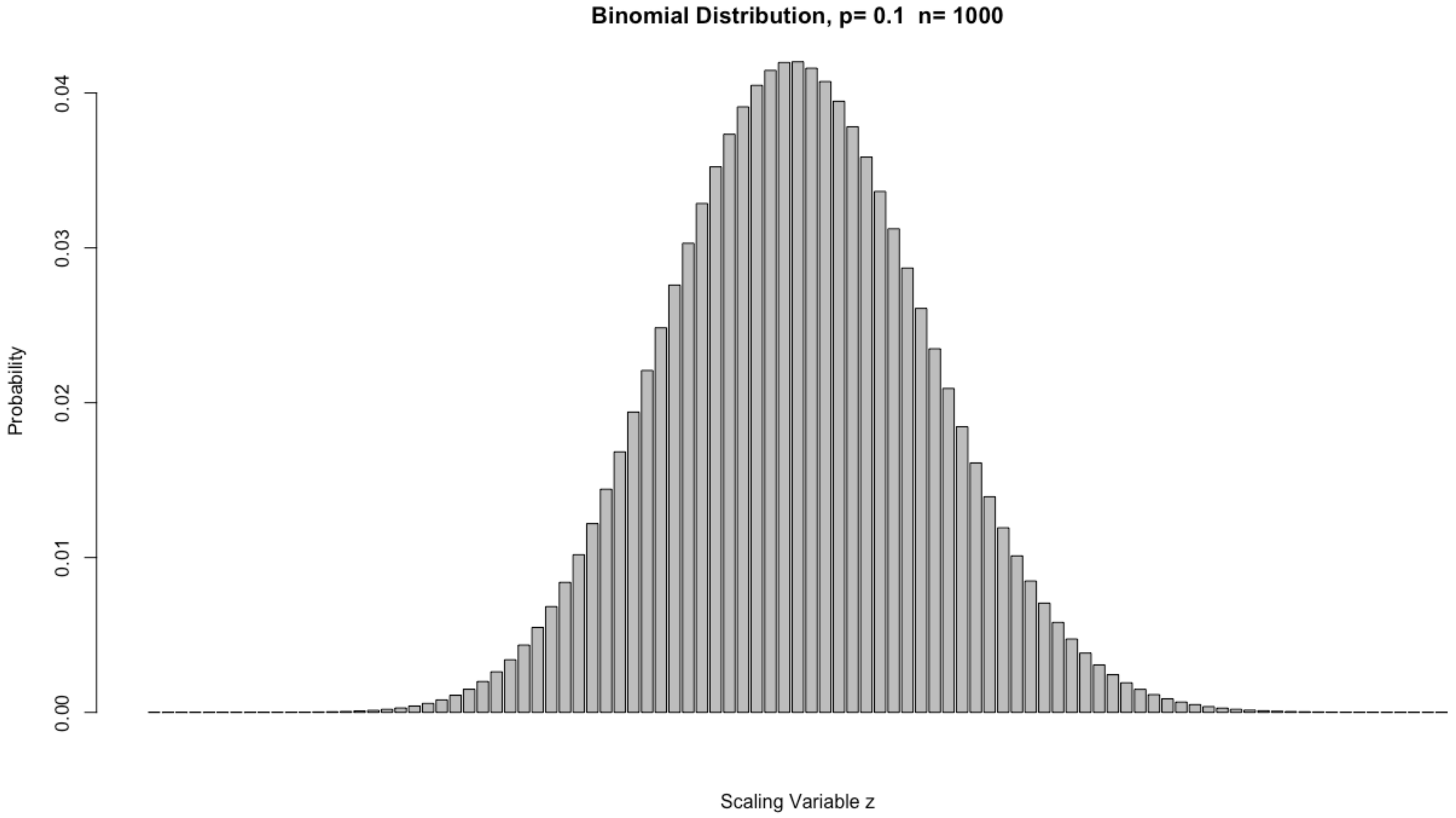
Binomial distribution

- Define the **scaling variable** $$z = z_k = \frac{k - np}{\sqrt{npq}}$$

```
nlist <- c(1,2,5,10,20,50,100,1000)
p     <- 0.1
zmax  <- 5

for (n in nlist) {
          k  <- 0:n
          z  <- (k - n*p)/sqrt(n*p*(1-p))
          zi <- (abs(z)<=zmax)
          f  <- dbinom(k,n,p)
          barplot(f[zi], xlab="Scaling Variable z", ylab="Probability",
          main=paste("Binomial Distribution, p=",p," n=",n))
          readline()
}
```

Scaled binomial distribution



Binomial distribution

- In terms of the scaling variable, the limiting distribution is **Gaussian**

$$z = z_k = \frac{k - np}{\sqrt{npq}}$$

$$\boxed{f(k; n, p) \approx f(z_k) = \frac{1}{\sqrt{2\pi}} e^{-z_k^2/2}}$$

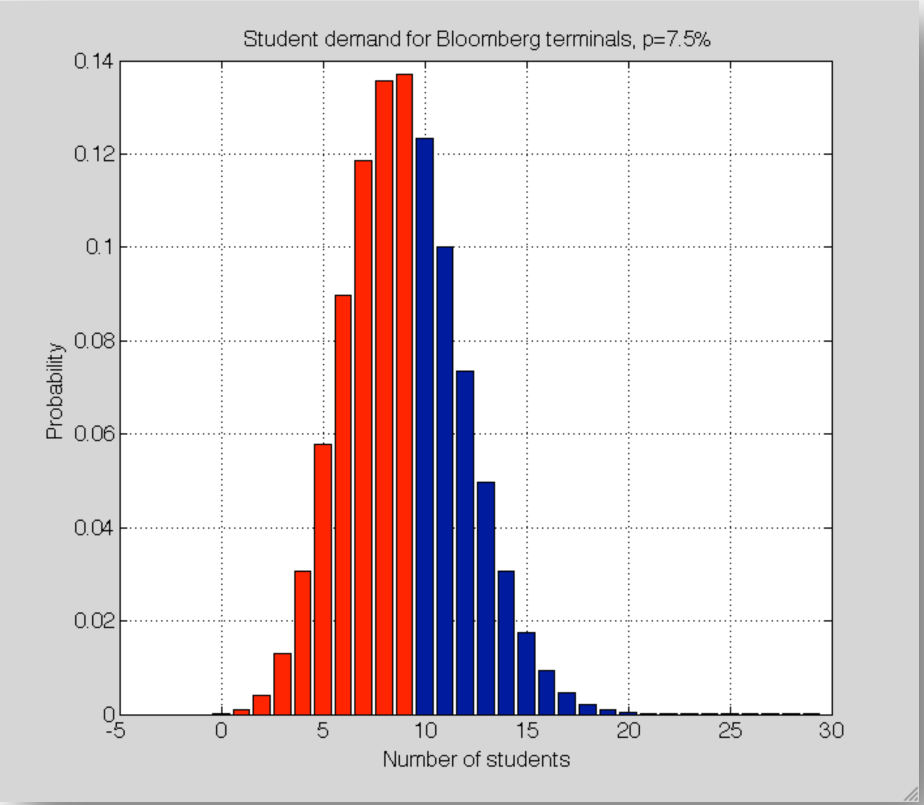
Example: Bloomberg terminals

- Parameters:
  - 120 students
  - 9 terminals
  - Demand independently with probability 7.5%

- Cumulative probability 58.75% that 9 or fewer students will want to use at same time.

```
n <- 120
k <- 0:9
p <- 0.075

sum(dbinom(k,n,p))
[1] 0.587475
```



Sums of random variables

Summary

- A set of individual random variables can be added to create a new random variable, their sum.

$$S_n = X_1 + X_2 + \cdots + X_n$$

- The full probability distribution can be computed using convolutions (which is hard), but we will usually need only the moments (which is easy).
- The **moments** of this new random variable can be easily computed using **linearity**.
- If the variables are **independent** and **identically distributed (IID)**, then the sum is **normally distributed** as $n \to \infty$.
- When the variables are ordered in **time**, the sequence of sums forms a **stochastic process**.

From moment generating functions to the CLT

Finance at MIT
Where ingenuity drives results

Characteristic function

- The **characteristic function** is defined so that its derivatives **generate** all the moments of the distribution

encoding all of the moments in one place

$$\tilde{f}(t) \equiv \mathrm{E}\left[e^{itX}\right], \quad \mathrm{E}\left[X^\ell\right] = (-i)^\ell \frac{\mathrm{d}^\ell}{\mathrm{d}t^\ell} \tilde{f}(t)\bigg|_{t=0}$$

- It is useful whenever the function and its derivatives can be computed in closed form. For the binomial distribution,

$$\begin{aligned}\tilde{f}(t) \equiv \mathrm{E}\left[e^{itX}\right] &= \sum_{k=0}^{\infty} e^{itk}\,\mathrm{Prob}(X=k) \\ &= \sum_{k=0}^{n} e^{itk} \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} \left(pe^{it}\right)^k q^{n-k} \\ &= \left(pe^{it}+q\right)^n\end{aligned}$$

Hence

$$\begin{aligned}\mathrm{E}\left[1\right] &= \tilde{f}(0) = (p+q)^n = 1, \\ \mathrm{E}\left[X\right] &= -i\tilde{f}'(0) = np, \\ \mathrm{E}\left[X^2\right] &= -\tilde{f}''(0) = npq + n^2p^2, \textit{ etc.}\end{aligned}$$

Characteristic function

"An example is the fat-tailed Cauchy distribution with a probability density function $p\left(x\right)=\frac{A}{\pi^2A^2+x^2}$*. Characteristic function of a Cauchy distribution is* $\tilde{p}\left(t\right)=e^{-\pi A|t|}$*. This is a generally well-behaved function, but its derivatives are not defined at* $t=0$*."*

- Continuous distributions: Fourier transform of density

$$\tilde{p}(t)\equiv\int_{-\infty}^{\infty}e^{itx}p(x)\,\mathrm{d}x=\sum_{\ell=0}^{\infty}\frac{(it)^\ell}{\ell!}\mathrm{E}\left[X^\ell\right],\qquad \mu_\ell=\mathrm{E}\left[X^\ell\right]=(-i)^\ell\frac{\mathrm{d}^\ell}{\mathrm{d}t^\ell}\tilde{p}(t)\bigg|_{t=0}$$

Especially useful for **sums of random variables**: $Y=X_1+X_2$,

- Density function of the sum = **convolution** of the individual functions
- Fourier transform of a convolution = **product** of the Fourier transforms

$$p(y)=\int p_1(x_1)p_2(y-x_1)\mathrm{d}x_1,$$

$$\tilde{p}(t)=\tilde{p}_1(t)\tilde{p}_2(t)$$

Normal distribution

- Gaussians are very special
  - The Fourier transform of a Gaussian is also a Gaussian

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \qquad \tilde{p}(t) = e^{-\sigma^2 t^2/2 + i\mu t}$$

Normal distribution

- The sum of *N* independent Gaussian random variables is itself a Gaussian RV with its own mean and variance:

$$
\begin{aligned}
\tilde{p}(t) &= \tilde{p}_1(t)\tilde{p}_2(t)\cdots\tilde{p}_N(t) \\
&= \exp\left[-\frac{\sigma_1^2 t^2}{2} + i\mu_1 t\right]\exp\left[-\frac{\sigma_2^2 t^2}{2} + i\mu_2 t\right]\cdots \\
&= \exp\left[-\frac{(\sigma_1^2 + \cdots + \sigma_N^2)t^2}{2} + i(\mu_1 + \cdots + \mu_N)t\right] \\
&= \exp\left[-\frac{\hat{\sigma}^2 t^2}{2} + i\hat{\mu}t\right],
\end{aligned}
$$

where

$$
\hat{\sigma}^2 = \sum \sigma_i^2, \quad \hat{\mu} = \sum \mu_i
$$

Scaling for Gaussians

- The sum of *N* independent, **identical** Gaussian random variables is a Gaussian RV with

$$\mu \to \hat{\mu} = N\mu,$$
$$\sigma \to \hat{\sigma} = \sqrt{N}\sigma$$

Cumulant expansion

$$\tilde{p}(t) \equiv \int_{-\infty}^{\infty} e^{itx} p(x)\, \mathrm{d}x = \sum_{\ell=0}^{\infty} \frac{(it)^\ell}{\ell!} \mathrm{E}\left[X^\ell\right], \qquad \mu_\ell = \mathrm{E}\left[X^\ell\right] = (-i)^\ell \frac{\mathrm{d}^\ell}{\mathrm{d}t^\ell} \tilde{p}(t)\bigg|_{t=0}$$

- Since Fourier transforms multiply, their logarithms **add**. Define the **cumulant** expansion,

$$\tilde{p}(t) \equiv \exp\left[\sum_{\ell=0}^{\infty} \frac{(it)^\ell}{\ell!} C_\ell\right], \qquad C_n = (-i)^n \frac{\mathrm{d}^n}{\mathrm{d}t^n} \log \tilde{p}(t)\bigg|_{t=0}$$

- The first few terms are

$$\begin{aligned}
C_1 &= \langle X \rangle, \\
C_2 &= \langle X^2 \rangle - \langle X \rangle^2, \\
C_3 &= \langle X^3 \rangle - 3\langle X \rangle \langle X^2 \rangle + 2\langle X \rangle^3, \\
C_4 &= \langle X^4 \rangle - 3\langle X^2 \rangle^2 - 4\langle X \rangle \langle X^3 \rangle + 12\langle X \rangle^2 \langle X^2 \rangle - 6\langle X \rangle^4
\end{aligned}$$

- For a Gaussian, all cumulants are exactly zero for $n > 2$.

Central limit theorem

- Consider the cumulants for a **sum of *N* independent, identically-distributed** random variables. Since cumulants add,

$$\hat{C}_n = NC_n,$$

Since Fourier transforms multiply, their logarithms add

- Then the dimensionless, normalized cumulants have a power-law dependence on *N:*

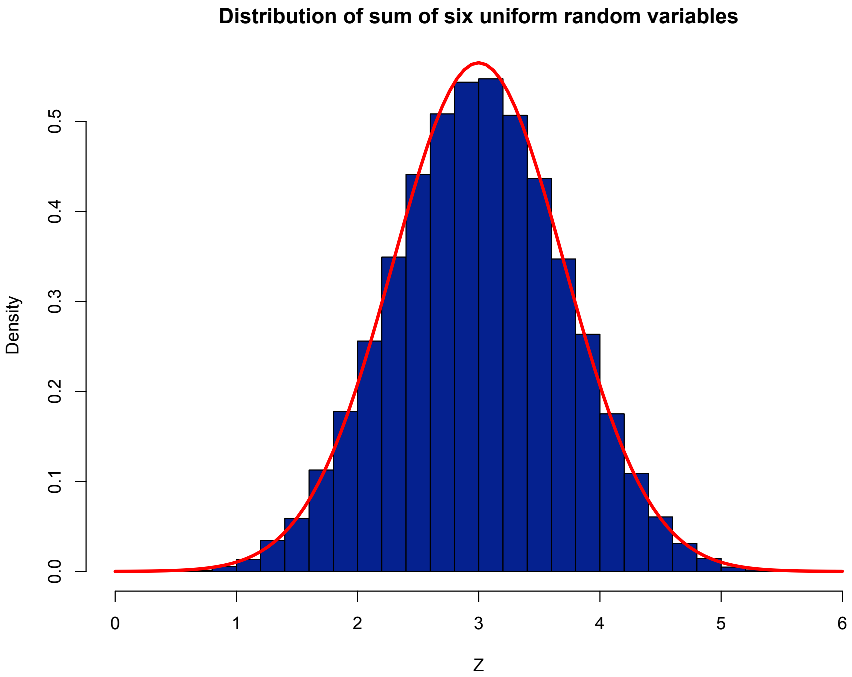
$$\frac{\hat{C}_n}{\hat{\sigma}^n} = \frac{1}{N^{n/2-1}}\left(\frac{C_n}{\sigma^n}\right)$$

- As $N \to \infty$, all cumulants vanish for $n > 2$, and the distribution for the sum **approaches a Gaussian.**

- However, this provides no information about the rate of convergence for different values of *x*; i.e., tails vs. center.

Six random variables: simulation

```
#Add 6 random variables drawn from uniform distribution on [0,1]
Z <- rowSums(matrix(runif(6e5),ncol=6))
hist(Z,probability=TRUE, breaks=30, col="#05218F",xlim=c(0,6),
 main="Distribution of sum of six uniform random variables")
curve(dnorm(x,mean=mean(Z),sd=sd(Z)),0,6,col="red",lwd=3,add=T)
```



Characteristic functions and the CLT

Summary

- The **characteristic function** of a probability distribution gives a compact formula for generating all of the moments.
- For continuous random variables, it is given by Fourier transforms, which have simple properties when applied to sums.
- Gaussians are special
  - The Fourier transform of a Gaussian is also a Gaussian.
  - A sum of Gaussian random variables is also a Gaussian random variable.
  - The central limit theorem (CLT) shows that Gaussians are **universal**: adding together a large number of **any** IID random variables approaches a Gaussian distribution*.

**Conditions apply. You mileage may vary.*