EC381/MN308
Probability and Some Statistics

Yannis Paschalidis

yannisp@bu.edu, http://ionia.bu.edu/

Dept. of Manufacturing Engineering

Dept. of Electrical and Computer Engineering

Center for Information and Systems Engineering

Lecture 18 - Outline

1. **Sums of Random Variables**
   a. **Motivation for limit theorems**
   b. **Transforms**
   c. **Central limit theorem**

Limit Theorems

- Objective: Given many joint random variables in an experiment:
  – Can we determine limits of functions of these random variables?
  – Example: average of many random variables
- Example: Collecting many independent samples for an experiment
  – Theoretical foundations of empirical statistics.

What are limit theorems?

- Limit theorems specify the probabilistic behavior of sums of random variables as $n \to \infty$
- Limits as in calculus, but now deal with a limit of a random sequence
  – What does convergence in this context mean?
- Will require restrictions on RVs so limits exist, such as:
  – Independent random variables
  – Uncorrelated random variables
  – Identical marginal CDFs/PDFs/PMFs
  – Identical means and/or variances
  – Other …

Motivating Example

- Bernoulli random variable X with parameter p
- Suppose we repeat the experiment independently, generating samples X_i, $i = 1, 2, \ldots, n$
- Define the derived random variable $Z_n = (X_1 + X_2 + \ldots + X_n)/n$
  – This is the sample average
- What happens to Z_n as $n \to \infty$?
  – Should be close to p, one hopes… make this precise!

Mean/Variance of Sum of RVs

- Consider $W_n = X_1 + X_2 + \ldots + X_n$

$$E[W_n] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \mu_i$$

$$Var(W_n) = \sum_{i=1}^{n} E\left[\left(\sum_{i=1}^{n}(X_i - \mu_i)\right)^2\right]$$

$$= E\left[\left(\sum_{i=1}^{n}(X_i - \mu_i)\right)\left(\sum_{i=1}^{n}(X_i - \mu_i)\right)\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)$$

$$= \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(X_i, X_j)$$

Average of n Random Variables

- Assume joint RVs $X_1, \ldots, X_n$ with finite expected values $\mu_1, \ldots, \mu_n$
- Define derived RV $Z_n = (X_1 + \ldots + X_n)/n$
- $E[Z_n] = (E[X_1] + \ldots + E[X_n])/n$
  - Linearity of $E[\bullet]$
- $E[Z_n] = (\mu_1 + \ldots + \mu_n)/n$
- Note: if all X_i had same mean μ, then $E[Z_n] = \mu$ for all n

Average of n RVs - 2

- Motivation: an experiment that generates RV X is repeated independently to generate samples X_i
  - Assume X_i has finite mean μ
  - Since X_i are generated by independent samples of same experiment, they are i.i.d. (independent and identically distributed)
- Definition: *Sample mean* Z_n is average of X_i

  $Z_n = (X_1 + \ldots + X_n)/n$
- What happens to Z_n as $n \to \infty$?

Average of n RVs - 3

- Z_n is a derived random variable

  Has mean $E[Z_n] = E[X_i] = \mu$
- Since Z_n is sum of independent random variables (uncorrelated would have been enough)

  $\text{Var}[Z_n] = \text{Var}[X_1/n] + \ldots + \text{Var}[X_n/n]$
- Note:

  $\text{Var}[X_1] = \text{Var}[X_i]$ for all i

  $\text{Var}[X_i/n] = \text{Var}[X_i]/(n^2)$ by scaling law

→ $\text{Var}[Z_n] = n\,\text{Var}[X_i]/(n^2) = \text{Var}[X]/n$

Average of n RVs - 4

- As $n \to \infty$, the variance of $Z_n \to 0$

  Less and less uncertainty in Z_n

  This implies that Z_n is very close to its average μ
- Note that the above results would hold even if the variables X_i were uncorrelated instead of independent

  All that is needed is $\text{Cov}[X_i, X_j] = 0$ if i differs from j
- Z_n is an estimate of μ, the expected value of X

  Given many samples, estimate has a very small error

Example: Quiz 7.1

- Let X be exponential with $\lambda=1$ ($E[X] = 1/\lambda=1$). Let Z_n denote the sample mean of n independent samples of X.

  How many samples are needed so that the variance of the sample mean is less than 0.01?

  We have $\text{Var}[X] = 1/\lambda^2 = 1$.

  Thus, $\text{Var}[Z_n] = \text{Var}[X]/n = 1/n$

  → Requires 101 samples since 100 samples gives exactly 0.01

A Useful Statistic: Moment Generating Function

- For continuous RVs, density $f_X(x)$ integrates to 1
  - It has a Fourier Transform!
  - It has a two-sided Laplace transform with a region of convergence
- Definition: The moment generating of a RV X is defined as

  $\Psi_X(s) = E[e^{sX}]$

$$\Psi_X(s) = E[e^{sX}] = \int_{-\infty}^{\infty} e^{sx} f_X(x)dx$$

Called $\phi_X(s)$ in Yates/Goodman text

Two-sided Laplace transform of $f_X(x)$

Can look them up in tables

Moment Generating Functions: Why would one care?

- Can get moments by differentiating the MGF (which is why it has that name)

$$\Psi_X(0) = 1$$
$$\frac{d}{ds}\Psi_X(s) = \int_{-\infty}^{\infty} \frac{d}{ds} e^{sx} f_X(x)dx = \int_{-\infty}^{\infty} x e^{sx} f_X(x)dx$$
$$\frac{d}{ds}\Psi_X(s)|_{s=0} = \int_{-\infty}^{\infty} x e^{0} f_X(x)dx = E[X]$$
$$\frac{d^2}{ds^2}\Psi_X(s)|_{s=0} = E[X^2]$$
$$\frac{d^n}{ds^n}\Psi_X(s)|_{s=0} = E[X^n]$$

A Most Useful Property

- X, Y independent jointly continuous RVs
- $Z = X + Y$

$$\Psi_Z(s) = E[e^{s(X+Y)}]$$
$$= E[e^{sX}]E[e^{sY}] \text{ independence}$$
$$= \Psi_X(s)\Psi_Y(s)$$

  – The Laplace (or Fourier) transform of a convolution is the product of the Laplace (or Fourier) transforms, as in signals and systems

A Particularly Useful Moment Generating Function: the Gaussian

- X is a standard Gaussian, zero mean, variance 1:

$$\Psi_X(s) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{sx} e^{-x^2/2} dx$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2+sx-s^2/2+s^2/2} dx$$
$$= e^{s^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-s)^2/2} dx$$
$$= e^{s^2/2}$$

THE MGF OF A GAUSSIAN PDF IS ITSELF GAUSSIAN!

Table 6.1 (p. 249)

Moment generating function for families of random variables.

Random Variable		PMF or PDF	MGF $\phi_X(s)$
Bernoulli (p)	$P_X(x) =$	$\begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & \text{otherwise} \end{cases}$	$1-p+pe^s$
Binomial (n, p)	$P_X(x) =$	$\binom{n}{x} p^x (1-p)^{n-x}$	$(1-p+pe^s)^n$
Geometric (p)	$P_X(x) =$	$\begin{cases} p(1-p)^{x-1} & x=1,2,\ldots \\ 0 & \text{otherwise} \end{cases}$	$\frac{pe^s}{1-(1-p)e^s}$
Pascal (k, p)	$P_X(x) =$	$\binom{x-1}{k-1} p^k (1-p)^{x-k}$	$(\frac{pe^s}{1-(1-p)e^s})^k$
Poisson (α)	$P_X(x) =$	$\begin{cases} \alpha^x e^{-\alpha}/x! & x=0,1,2,\ldots \\ 0 & \text{otherwise} \end{cases}$	$e^{\alpha(e^s-1)}$
Disc. Uniform (k, l)	$P_X(x) =$	$\begin{cases} \frac{1}{l-k+1} & x=k,k+1,\ldots,l \\ 0 & \text{otherwise} \end{cases}$	$\frac{e^{sk}-e^{s(l+1)}}{1-e^s}$
Constant (a)	$f_X(x) =$	$\delta(x-a)$	e^{sa}
Uniform (a, b)	$f_X(x) =$	$\begin{cases} \frac{1}{b-a} & a<x<b \\ 0 & \text{otherwise} \end{cases}$	$\frac{e^{bs}-e^{as}}{s(b-a)}$
Exponential (λ)	$f_X(x) =$	$\begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$	$\frac{\lambda}{\lambda-s}$
Erlang (n, λ)	$f_X(x) =$	$\begin{cases} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$	$(\frac{\lambda}{\lambda-s})^n$
Gaussian (μ, σ)	$f_X(x) =$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$	$e^{s\mu+s^2\sigma^2/2}$

Random Sum of i.i.d. RVs

- $R=X_1+\cdots+X_N$, X_i are i.i.d., N is random

$$E[R] = E[E[R|N=n]] = E[nE[X]] = E[N]E[X]$$

$$E[R^2] = E[E[R^2|N=n]]$$
$$= E[E[R^2 - (E[R|N=n])^2|N=n]] + E[(E[R|N=n])^2|N=n]]$$
$$= E[Var(R|N=n)] + E[n^2(E[X])^2]$$
$$= E[nVar(X)] + E[N^2](E[X])^2$$
$$= E[N]Var(X) + E[N^2](E[X])^2$$
$$Var(R) = E[R^2] - (E[N]E[X])^2 = E[N]Var(X) + (E[X])^2 Var(N)$$

$$E[e^{sR}] = E[E[e^{s(X_1+\cdots+X_N)}|N=n]]$$
$$= E[(\Psi_X(s))^n]$$
$$= E[e^{n \ln \Psi_X(s)}] = \Psi_N(\ln \Psi_X(s))$$

The Central Limit Theorem

- SLLN says $Z_n = (X_1 + \ldots + X_n)/n$ converges to μ with probability one
  – Implies the CDF of Z_n converges to a unit step at μ
  – Can we say more?
- Assume the X_i have finite variance σ_X^2.

  Then, $Z_n - \mu$ is a zero-mean random variable with variance σ_X^2/n

$$Y_n = \frac{\sqrt{n}(Z_n - \mu)}{\sigma_X}$$

  – Normalize:
    - Variance 1, mean zero for all n.
    - What is the CDF of Y_n?

CLT - 2

- Central Limit Theorem: Given iid RVs with finite mean μ, finite variance σ_X^2, the CDF of

$$Y_n = \frac{\sqrt{n}((X_1 + \ldots + X_n)/n - \mu)}{\sigma_X}$$

converges to $\Phi(\bullet)$, the CDF of a unit Gaussian RV

➔ For each real number y,

$$\lim_{n \to \infty} P[\frac{\sqrt{n}((X_1 + \ldots + X_n)/n - \mu)}{\sigma_X} \leq y] = \Phi(y)$$

CLT – 3

- The CLT allows us to compute probabilities of Z_n for finite, but large n
  - Probabilities referring to the differences between Z_n and μ can be approximated by the Gaussian CDF
  - Note: this does not say the difference between Z_n and μ is a Gaussian
  - Nevertheless, this explains the popularity and importance of Gaussian models

CLT - 4

- Computation: rescaling, this can be interpreted as approximating the CDF of the sum (X_1 + … + X_n) by that of a Gaussian, N($n\mu$, $n\sigma_X^2$)

$$P[(X_1 + \ldots + X_n) \leq a] \approx \Phi(\frac{a - n\mu}{\sqrt{n}\sigma_X})$$

  - Good approximation for $|a - n\mu| < 3\, n^{1/2}\, \sigma_X$
  - Not so good for outliers ("tails" of the distribution")

Using the CLT

- Quiz 6.6: X milliseconds for each disk access time, uniformly distributed in [0,12]. Assume one must access disk 12 times independently.
  - $T = X_1 + \ldots + X_{12}$
  - $E[T]$ = 12 $E[X]$ = 12 * 6 = 72 msec
  - Var [T] = 12 Var[X] = 12 * 12^2/12 = 144 msec^2
  - σ_T = 12 msec
  - BUT what if we want to know $P(T > 75$ msec) = ???

Turn to the CLT: T is sum of iid RVs, finite variances ➔ CDF of T is approximately N(72,144)

$P(T > 75) \rightarrow$ Q(3/12) = Q(0.25)

$P(T < 48) \rightarrow$ Q(24/12) = Q(2)

To do this exactly would require 12 convolutions!

Using CLT again

- Modem transmits 10^4 bits
  - Each bit 0 or 1, i.i.d., p = 0.5
  - Estimate P(number of 1's > 5100)
  - Estimate P(number of 1's ∈ (4800,5100))

T = number of 1's

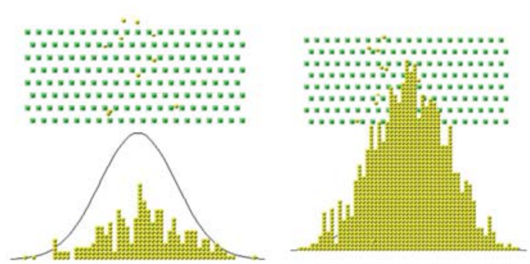
$E[T]$ = 5000; Var[T] = 10^4*Var[Bernoulli(0.5)] = 2500, so standard deviation is 50

$P(T > 5100) \rightarrow Q(100/50) = Q(2)$

$P(T < 4800) \rightarrow Q(4)$

$P(T \in (4800,5100)) \rightarrow 1 - Q(2) - Q(4)$

CLT Demonstration



Proving the CLT

- Define the derived random variable:

$$Y_n = \frac{X_1 - \mu}{\sqrt{n}\sigma_X} + \ldots + \frac{X_n - \mu}{\sqrt{n}\sigma_X}$$

- New random variables W_i are defined as

$$W_i = \frac{X_i - \mu}{\sqrt{n}\sigma_X}$$

The W_i are iid, zero-mean, variance $1/n$

Proving the CLT - 2

- MGF of *W:*

$$\Psi_W(s) = E[e^{sW}] = E[e^{\frac{s}{\sqrt{n}\sigma_X}(X-\mu)}]$$
$$= e^{-\frac{s}{\sqrt{n}\sigma_X}\mu}\Psi_X(\frac{s}{\sqrt{n}\sigma_X})$$

- MGF of Y_n

$$\Psi_{Y_n}(s) = \Psi_W(s)^n \ = e^{-n\frac{s}{\sqrt{n}\sigma_X}\mu}\Psi_X(\frac{s}{\sqrt{n}\sigma_X})^n$$

CLT Proof - 3

- Let X_i be zero mean (wlog, simplifies algebra)

$$\Psi_{Y_n}(s) = \Psi_W(s)^n = \Psi_X(\frac{s}{\sqrt{n}\sigma_X})^n$$

- Now, note that $\Psi_X(s)$ has a Taylor series expansion around zero, as

$$\Psi_X(s) = \Psi_X(0) + s\frac{d}{d\omega}\Psi_X(s)|_{\omega=0} + \frac{s^2}{2!}\frac{d^2}{ds^2}\Psi_X(s)|_{s=0} + O(s^3)$$
$$= 1 + \frac{s^2}{2}\frac{d^2}{ds^2}\Psi_X(s)|_{\omega=0} + O(s^3) \ (zero\ mean)$$
$$= 1 + \frac{s^2}{2}\sigma_X^2 + O(s^3) \ (moment\ generation)$$
$$= e^{\frac{s^2}{2}\sigma_X^2} + O(s^3)$$

CLT Proof - 4

- Now, substitute back:

$$\Psi_{Y_n}(s) = \Psi_X(\frac{s}{\sqrt{n}\sigma_X})^n$$
$$= (e^{\frac{s^2}{n\sigma_X^2}\frac{\sigma_X^2}{2}} + O(\frac{s^3}{n^{3/2}\sigma_X^3}))^n$$
$$= e^{s^2/2}[1 + nO(\frac{s^3}{n^{3/2}\sigma_X^3})e^{\frac{s^2}{2n}} + \ldots + (O(\frac{s^3}{n^{3/2}\sigma_X^3})e^{\frac{s^2}{2n}})^n]$$

- As $n \rightarrow \infty$, all terms but the first go to zero for each fixed s!

$$\lim_{n\to\infty} \Psi_{Y_n}(s) = e^{s^2/2}$$

  - The MGF of the scaled difference from the mean, Y_n, converges to the Gaussian MGF, leads to the CLT