# Sums of Random Variables and The Law of Large Numbers

Mark Huiskes, LIACS
mark.huiskes@liacs.nl

# Introduction

- Today we look at sums of independent random variables and after that at averages which are very much related to the sums
- A lot of estimators, for instance sample means/averages and also many estimators occurring in pattern recognition, are ultimately sums of independent random variables, so it is important to be able to calculate their distributions and derived properties.
- Let's first look at the sum of two independent variables in the discrete case: $Z = X + Y$. We already know: $E(Z) = E(X) + E(Y)$, and $V(Z) = V(X) + V(Y)$. [But how to compute the full distribution].

# Convolutions

- Independent random variables X and Y with distribution functions m_X(x) and m_Y(y). What is m_Z(z) [the distribution function] of Z = X + Y?
- Look at the event {Z = z} in terms of the outcomes of X and Y.
- Let's look at an example. Suppose m_X = ( 1 2\\ ¼ ¾) and m_Y = ( 0 1 2// ½ ¼ ¼). What is m_Z, with Z = X + Y?

- General pattern: {Z = z} = \cup ({X = k} \cap {Y = z – k})
- [This is a union of disjoint events, so the probability of P(Z = z) is the sum of the probabilities of these intersections:

$$P(Z = z) = \sum_{k=-\infty}^{\infty} P(X = k, Y = z - k) = \sum_{k=-\infty}^{\infty} P(X = k) \cdot P(Y = z - k)$$

# Convolutions (continued)

- This distribution function is called the convolution of m_X and m_Y and is written as: m_Z = m_X * m_Y
- [further explain in terms of example]
- Explain that you can repeat this procedure indefinitely to compute the distribution of any sum of variables. Tedious job so generally done by computer.

# The Sum of Continuous Variables

- Let X and Y be two continuous rv's with density functions f(x) and g(y), resp. Then the convolution f * g of f and g is the function given by

$$(f * g)(z) = \int_{-\infty}^{\infty} f(x)g(z-x)\,dx = \int_{-\infty}^{\infty} g(y)f(z-y)\,dy$$

- If X and Y are independent then the density of their sum is the convolution of their densities.

$$\approx f_Z(z) = \cup_x (X = x \cap Y = (z-x))$$

# The Sum of Two Uniform Variables

- X ~ U[0,1], Y~U[0,1]
- Z = X + Y by convolution:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx = \int_0^1 f_Y(z-x)\,dx$$

- z is a fixed number; unless 0 <= z – x <= 1 the integrand is zero, i.e. unless z – 1 <= x <= z [i.e. x between z-1 and z].
- If z <= 1, then

$$f_Z(z) = \int_0^1 f_Y(z-x)\,dx = \int_0^z dx = z$$

- If 1 <= z <= 2:

$$f_Z(z) = \int_0^1 f_Y(z-x)\,dx = \int_{z-1}^1 dx = 2-z$$

[Write down density formula (0 elsewhere), and draw graph]

# Sums of Normal Random Variables

- The convolution of two normal densities with parameters mu1, sig1^2 and mu2, sig2^2 is again a normal density with parameters mu1+mu2 and sig1^2 + sig2^2.
- Or: X~N(mu1, sig1^2), Y~N(mu2, sig2^2)
  Z=X+Y~N(mu1+mu2, sig1^2+sig2^2)

# [What you should know of Ch 7]

- Be able to compute the density of a sum of independent discrete variables.
- Understand the expression for the convolution of continuous variables
- Know how to compute the density of sums of independent random normal variables with given parameters.

# Ch 8: The Law of Large Numbers

- The law of large numbers is a fundamental theorem of probability. It describes the convergence of sample averages to the true mean when you take more and more samples. It's mainly useful if you get to the more advanced statistical theory, but it's good to have seen it already. Next week we also discuss the central limit theorem that shows that sums of random variables tend to get a normal distribution if you sum many independent variables.
- Let's first take a look at what averages are really.

# Averages

- Consider an independent trial process: remember this is a sequence of random variables X1, X2, X3 ,… each with the same distribution. Assume E(X_i) = mu and V(X_i) = sig^2 (for all i.)
- S_n = X_1 + X_2 + … + X_n
- E(S_n) = n mu and Var(S_n) = n sig^2 [Note that S_n not equal to n.X!!! Then we would have had n^2 sig^2]
- A_n = S_n / n
- [Average or Mean is a random variable!!!!]
- E(A_n) = mu
- V(A_n) = sig / n, so D(A_n) = sig / sqrt(n): standard deviation of the mean, also known as standard error

# Law of Large Numbers

- LaIn states that average tends to the mean when n to infinity with high certainty.
- From the standard error we already see that when n gets large we can expect small deviations.
- There are strong versions of the law of large numbers which are harder to proof that show that the sample mean really converges to the true mean.
- The book proofs the so-called weak law of large numbers, which can be proofed with a simple inequality, but which shows only that the probability of deviation from the mean are unlikely.

# Chebyshev Inequality

- X a discrete random variable with expectation E(X) = mu, and variance V(X). Let eps >0 be any positive real number. Then

$$P(|X - \mu| \geq \varepsilon) <= \frac{V(X)}{\varepsilon^2}$$

- The inequality captures a relation between the probability that the rv differs a certain distance from the mean and the variance.

- Important consequence: take \eps = k \sig, then

$$P(|X - \mu| \geq k\sigma) <= \frac{1}{k^2}$$

- Note this inequality works for any distribution!!! This is also why the bound is not always that good. If you have more information on the distribution you can find better bounds.

# Assignment 1

- X: #heads in 100 coin tosses. EX = 50 D(X): sqrt(npq) = 5.
- What does Chebyshev tell us about the probability that the number of heads deviates by more than 3 standard deviations from the mean?

P( X<5 or X>45) <= 1/9

# Law of Large Numbers (Discrete Case)

- Let X1, X2, …, Xn be an independent trials process, with finite expected value mu = E(X_j) and finite variance \sig^2 = V(X_j). Let Sn = X1+ … + Xn. Then for any \eps >0

$$P(|\frac{S_n}{n} - \mu| \geq \varepsilon) \to 0, \ \text{as}\, n \to \infty$$

- Or, equivalently:

$$P(|\frac{S_n}{n} - \mu| < \varepsilon) \to 1$$

- So, if we base an average on a sufficient number of outcomes it will be very close to the mean with a high degree of certainty.
- This provides a justification for the frequency interpretation of probability.

# Proof

- Directly consequence of applying Chebyshev to average

# Law of Large Numbers for Continuous Variables

- Chebyshev is exactly the same.

$$P(|X - \mu| \geq \varepsilon) <= \frac{V(X)}{\varepsilon^2}$$

- Also the Law of Large Numbers is the same (proof completely analogous).

- Discuss the Uniform Case (p317)

- Make assignment 10.