Week 7 – Transformation of RVs, Law of large numbers

Juan Pablo Lewinger

Last Class

- Random vector = multiple random variables defined on the same probability space.
- Discrete (have joint pmf) and continuous (have joint pdf) random vectors
- LOTUS to compute the expectation of a transformed random vector.
- Independence of random variables.
- Covariance and correlation.

Distribution of a sum of discrete RVs

(X, Y) a random vector. What is the distribution of $Z = X + Y$?

- If (X, Y) is discrete then $P_Z(z) = \sum_{\substack{x_i + y_i = z \\ (x_i, y_i) \in R_{X,Y}}} P_{X,Y}(x_i, y_i)$
- Example: Let X and Y be independent random variables with common pmf given by $P_X(0) = \frac{1}{4}$, $P_X(1) = \frac{1}{2}$, $P_X(2) = \frac{1}{4}$. Find the pmf of $Z = X + Y$.

Distribution of a sum of continuous RVs

(X, Y) a random vector and $Z = X + Y$

- If (X, Y) is continuous then $f_Z(z) = \int_{-\infty}^{+\infty} f_{X,Y}(x, z - x)\, dx = \int_{-\infty}^{+\infty} f_{X,Y}(z - y, y)\, dy$
- Example: $X \sim Exp(\lambda)$, $Y \sim U[0, 1]$, X and Y independent. Find the cdf and the pdf of $Z = X + Y$.

Bivariate normal distribution

Two random variables X and Y are said to have a bivariate normal distribution if their joint pdf is given by:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}$$

$\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} = E\begin{pmatrix} X \\ Y \end{pmatrix}$ and

$\boldsymbol{\Sigma} = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$ is the variance covariance-matrix.

where $\rho \in (-1,1)$, $\sigma_X > 0, \sigma_Y > 0$.

$(X,Y) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Sums of normals

- $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ are independent then $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

- $(X, Y) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $X + Y \sim N(\mu_X + \mu_Y,\ \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y)$

- If $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ but not bivariate normal then $X + Y$ are not necessarily normal.

  Example: $X \sim N(\mu, \sigma^2)$, $Y = -X$, $Y \sim N(\mu, \sigma^2)$ but $X + Y = 0$, a constant random variable

Sum of independent exponentials

$X_1, \ldots, X_n \overset{iid}{\sim} Exp(\lambda)$ then $Z = X_1 + \ldots + X_n \sim Gamma(n, \lambda)$

A continuous random variable has a gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$ if its probability density function is given by:

$$f(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} I_{[0,+\infty)}(x)$$

where $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t}\, dt$ for $\alpha > 0$ is the gamma function.

The gamma function is a generalization of the factorial: $\Gamma[\alpha + 1] = \alpha\Gamma[\alpha]$ and $\Gamma[n + 1] = n!$ for $n = 0, 1, 2, \ldots$

Maximum and minimum

$X_1, \ldots, X_n \sim F(x)$ iid random variables and $U = min(X_1, \ldots, X_n)$, $V = max(X_1, \ldots, X_n)$

Then:

$$F_V(v) = F(v)^n$$

(if continuous with density $f(x)$, then $f_V(v) = nf(v)F(v)^{n-1}$

$$F_U(u) = 1 - (1 - F(u))^n$$

(if continuous with density $f(x)$, then $f_U(u) = nf(u)(1 - F(u))^{n-1}$

Averaging reduces variability

$X_1, X_2 \sim F(x)$ independent

$Var(X_1) = Var(X_2) = \sigma^2$

$Var(\frac{X_1+X_2}{2}) = \frac{1}{2^2}(Var(X_1) + Var(X_2)) = \frac{1}{2^2}(\sigma^2 + \sigma^2) = \frac{\sigma^2}{2}$

Averaging reduces variability

$X_1, X_2, \ldots$ i.i.d with expectation μ and variance σ^2

$\overline{X_n} = \frac{X_1 + \ldots + X_n}{n}$

$$E[\overline{X_n}] = \frac{1}{n}(E[X_1] + \ldots + E[X_n]) = \frac{1}{n}(\underbrace{\mu + \ldots + \mu}_{n \text{ times}}) = \mu$$

$$Var[\overline{X_n}] = \frac{1}{n^2}(Var[X_1] + \ldots + Var[X_n]) = \frac{1}{n^2}(\underbrace{\sigma^2 + \ldots + \sigma^2}_{n \text{ times}}) = \frac{\sigma^2}{n}$$

Markov's inequality

If $U \geq 0$ is a random variable with finite expectation, then, for every $t > 0$:

$$P(U \geq t) \leq \frac{E[U]}{t}$$

Intuition:

- Let U be the income of a random selected individual from a population. Taking $t = 2E[U]$, Markov's inequality says that $P(U \geq 2E[U]) \leq \frac{1}{2}$. It is impossible for more than half of the population to make at least twice the average income.
- This has to be true because if more than half of the population make more than twice the average income, the average income would have to be higher!
- Taking $t = 3E[U]$, we get $P(U \geq 3E[U]) \leq \frac{1}{3}$; It is impossible for more than one third of the population to make at least 3 times the average income.

Markov's inequality

Proof

$U \geq 0$ a random variable with finite expectation, $t > 0$:

Define the new random variable $U_t = \begin{cases} 0 & \text{if } U < t \\ t & \text{if } U \geq t \end{cases}$

- U_t is discrete random variable taking only values 0 and t. So, $E[U_t] = tP(U \geq t)$.
- Clearly $U \geq U_t$
- Expectations preserve inequalities, so, $E[U] \geq E[U_t] = tP(U \geq t) \implies P(U \geq t) \leq \frac{E[U]}{t}$

Chebyshev's inequality

If X is a random variable with mean μ and variance σ^2, then, for every $t > 0$:

- $P\left(|X - E[X]| \geq t\right) \leq \frac{Var[X]}{t^2}$

    or equivalently

- $P\left(|X - E[X]| < t\right) \geq 1 - \frac{Var[X]}{t^2}$

    or equivalently

- $P\left(\frac{|X - E[X]|}{\sigma} \geq t\right) \leq \frac{1}{t^2}$

Chebyshev's inequality

Proof

X a random variable with mean μ and variance σ^2, $t > 0$

- Consider the non-negative random variable $U = (X - \mu)^2$
- Apply Markov's inequality to U, with t^2: $P(U \geq t^2) = P((X-\mu)^2 \geq t^2) \leq \frac{E[(X-\mu)^2]}{t^2} = \frac{Var(X)}{t^2}$
- Noticing that $\{|X-\mu| > t\} = \{(X-\mu)^2 > t^2\}$, we get Chebishev's inequality:

$$P(|X-\mu| \geq t) \leq \frac{Var(X)}{t^2}$$

Chebyshev’s inequality consequences

Any random variable with finite variance (continuous, discrete, mixed), regardless of its distribution, has most of its probability mass concentrated within a few standard deviations of its mean.

Taking $t = k\sigma, k = 2, 3, 4$ and applying Chebishev’s inequality:

$$P\left(|X - E[X]| \geq k\sigma\right) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

$$P\left(|X - E[X]| \leq k\sigma\right) \geq 1 - \frac{1}{k^2}$$

For $k = 2, 3, 4$ the bound are $3/4$, $8/9$, and $15/16$ respectively

Weak Law of large numbers (WLLN)

$X_1, X_2, \ldots$ i.i.d with finite expectation μ then:

$$\lim_{n \to \infty} P(|\overline{X_n} - \mu| > \epsilon) = 0$$

This form of convergence of random variables is called convergence in probability and is denoted as $\overline{X_n} \xrightarrow{P} \mu$

Equivalently:

$$\lim_{n \to \infty} P(|\overline{X_n} - \mu| \leq \epsilon) = 1$$

The sample mean converges to the true mean

(Weak) Law of large numbers

Proof

$X_1, X_2, \ldots$ i.i.d with finite expectation μ.

- To make the proof easier, we will also assume that $\sigma^2 < +\infty$
- But the for the WLLN to hold only the expectation needs to exist. The variance may be infinite.
- Let $\epsilon > 0$. Apply Chebyshev's inequality to $\overline{X}_n$:

$$P\left(|\overline{X}_n - \mu| \geq \epsilon\right) \leq \frac{Var(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

This implies:

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \epsilon) = 0$$

Strong Law of large numbers (SLLN)

$X_1, X_2, \ldots$ i.i.d with finite expectation μ.

$$P(\lim_{n \to \infty} \overline{X_n} = \mu) = 1$$

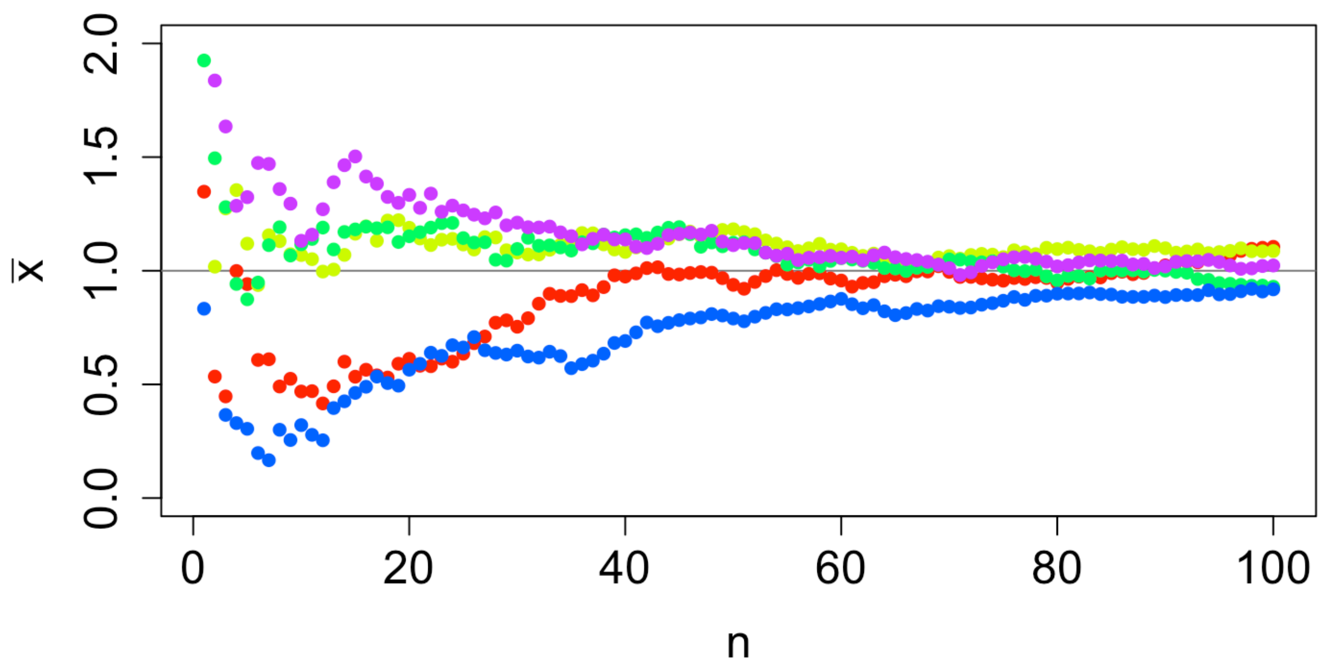
This form of convergence of random variables is called convergence with probability 1 and is denoted as $\overline{X_n} \xrightarrow{\text{w.p. } 1} \mu$ (also called almost sure convergence)

The strong law is stronger than weak law because convergence with probability 1 is a stronger concept than convergence in probability: convergence with probability 1 $\implies$ convergence in probability (but not the other way around)

Law of large numbers

$X_1, X_2, \ldots, X_n \sim N(1, 1)$

Five realizations of $\overline{X}_n$ as a function of n



Next two weeks

- Next week (Oct 21): take home midterm, no class
- For class Oct 30 read:
  - IPS 7.1.2
  - FSR 15.1-15.2