Probability review

Alejandro Ribeiro
Dept. of Electrical and Systems Engineering
University of Pennsylvania
aribeiro@seas.upenn.edu
http://www.seas.upenn.edu/users/~aribeiro/

August 31, 2015

Joint cdf

- Want to study problems with more than one RV. Say, e.g., X and Y
- Probability distributions of X and Y are not sufficient
  - $\Rightarrow$ Joint probability distribution of (X, Y). Joint cdf defined as

$$F_{XY}(x, y) = \text{P}\left[X \leq x, Y \leq y\right]$$

- If X, Y clear from context omit subindex to write $F_{XY}(x, y) = F(x, y)$
- Can write $F_X(x)$ by considering all possible values of Y

$$F_X(x) = \text{P}\left[X \leq x\right] = \text{P}\left[X \leq x, Y \leq \infty\right] = F_{XY}(x, \infty)$$

- Likewise $\Rightarrow F_Y(y) = F_{XY}(\infty, y)$
- In this context $F_X(x)$ and $F_Y(y)$ are called marginal cdfs

Joint pmf

- Discrete RVs X with possible values $\mathcal{X} := \{x_1, x_2, \ldots\}$ and Y with possible values $\mathcal{Y} := \{y_1, y_2, \ldots\}$
- Joint pmf of (X, Y) defined as

$$p_{XY}(x, y) = \text{P}\left[X = x, Y = y\right]$$

- Possible values (x, y) are elements of the Cartesian product $\mathcal{X} \times \mathcal{Y}$
  - (x_1, y_1), (x_1, y_2), ..., (x_2, y_1), (x_2, y_2), ..., (x_3, y_1), (x_3, y_2), ...
- $p_X(x)$ obtained by summing over all possible values of Y

$$p_X(x) = \text{P}\left[X = x\right] = \sum_{y \in \mathcal{Y}} \text{P}\left[X = x, Y = y\right] = \sum_{y \in \mathcal{Y}} p_{XY}(x, y)$$

- Likewise $\Rightarrow p_Y(y) = \sum_{x \in \mathcal{X}} p_{XY}(x, y)$
- Marginal pmfs

Joint pdf

- Continuous variables X, Y. Arbitrary sets $\mathcal{A} \in \mathbb{R}^2$
- Joint pdf is a function $f_{XY}(x,y) : \mathbb{R}^2 \to \mathbb{R}^+$ such that

$$\mathrm{P}\left[(X,Y) \in \mathcal{A}\right] = \iint_{\mathcal{A}} f_{XY}(x,y)\, dxdy$$

- Marginalization. There are two ways of writing $\mathrm{P}\left[X \in \mathcal{X}\right]$

$$\mathrm{P}\left[X \in \mathcal{X}\right] = \mathrm{P}\left[X \in \mathcal{X}, Y \in \mathbb{R}\right] = \int_{X \in \mathcal{X}} \int_{-\infty}^{+\infty} f_{XY}(x,y)\, dy\, dx$$

- From the definition of $f_X(x)$ $\Rightarrow \mathrm{P}\left[X \in \mathcal{X}\right] = \int_{X \in \mathcal{X}} f_X(x)\, dx$
- Lipstick on a pig (same thing written differently is still same thing)

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\, dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\, dx$$

Example

- Draw two Bernoulli RVs B_1, B_2 with the same parameter p
- Define $X = B_1$ and $Y = B_1 + B_2$
- The probability distribution of X is

$$p_X(0) = 1 - p, \quad p_X(1) = p$$

- Probability distribution of Y is

$$p_Y(0) = (1-p)^2, \quad p_X(1) = 2p(1-p), \quad p_X(2) = p^2$$

- Joint probability distribution of X and Y

$$\begin{array}{lll} p_{XY}(0,0) = (1-p)^2, & p_{XY}(0,1) = p(1-p), & p_{XY}(0,2) = 0 \\ p_{XY}(1,0) = 0, & p_{XY}(1,1) = p(1-p), & p_{XY}(1,2) = p^2 \end{array}$$

Random vectors

- For convenience arrange RVs in a vector.
- Prob. distribution of vector is joint distribution of its components
- Consider, e.g., two RVs X and Y. Random vector is $\mathbf{X} = [X, Y]^T$
- If X and Y are discrete, vector variable $\mathbf{X}$ is discrete with pmf

$$p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{X}}\left([x, y]^T\right) = p_{XY}(x, y)$$

- If X, Y continuous, $\mathbf{X}$ continuous

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}\left([x, y]^T\right) = f_{XY}(x, y)$$

- Vector cdf is $\Rightarrow F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}\left([x, y]^T\right) = F_{XY}(x, y))$
- In general, can define n-dimensional RVs $\mathbf{X} := [X_1, X_2, \ldots, X_n]^T$
- Just a matter of notation

Joint expectations

- RVs X and Y and function $g(X, Y)$. Function $g(X, Y)$ also a RV
- Expected value of $g(X, Y)$ when X and Y discrete can be written as

$$\mathbb{E}\left[g(X, Y)\right] = \sum_{x,y:p_{XY}(x,y)>0} g(x, y) p_{XY}(x, y)$$

- When X and Y are continuous

$$\mathbb{E}\left[g(X, Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\, dxdy$$

- Can have more than two RVs. Can use vector notation

Example

- Linear transformation of a vector RV $\mathbf{X} \in \mathbb{R}^n$: $g(\mathbf{X}) = \mathbf{a}^T\mathbf{X}$

$$\Rightarrow \mathbb{E}\left[\mathbf{a}^T\mathbf{X}\right] = \int_{\mathbb{R}^n} \mathbf{a}^T\mathbf{X} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

Expected value of a sum of random variables

- Expected value of the sum of two RVs,

$$\mathbb{E}[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f_{XY}(x,y)\,dxdy$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x\,f_{XY}(x,y)\,dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}y\,f_{XY}(x,y)\,dxdy$$

- Remove x (y) from innermost integral in first (second) summand

$$\mathbb{E}[X+Y] = \int_{-\infty}^{\infty}x\int_{-\infty}^{\infty}f_{XY}(x,y)\,dy\,dx + \int_{-\infty}^{\infty}y\int_{-\infty}^{\infty}f_{XY}(x,y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty}xf_X(x)\,dx + \int_{-\infty}^{\infty}yf_Y(y)\,dy$$
$$= \mathbb{E}[X] + \mathbb{E}[Y]$$

- Used marginal expressions

- Expectation $\leftrightarrow$ summation $\Rightarrow \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Expected value is a linear operator

- Combining with earlier result $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ proves that

$$\mathbb{E}[a_x X + a_y Y + b] = a_x \mathbb{E}[X] + a_y \mathbb{E}[Y] + b$$

- Better yet, using vector notation (with $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^n$, b a scalar)

$$\mathbb{E}\left[\mathbf{a}^T \mathbf{X} + b\right] = \mathbf{a}^T \mathbb{E}[\mathbf{X}] + b$$

- Also, if $\mathbf{A}$ is an $m \times n$ matrix with rows $\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T$ and $\mathbf{b} \in \mathbb{R}^m$ a vector with elements $b_1, \ldots, b_m$ we can write

$$\mathbb{E}\left[\mathbf{A}^T \mathbf{X} + \mathbf{b}\right] = \begin{pmatrix} \mathbb{E}\left[\mathbf{a}_1^T \mathbf{X} + b_1\right] \\ \mathbb{E}\left[\mathbf{a}_2^T \mathbf{X} + b_2\right] \\ \vdots \\ \mathbb{E}\left[\mathbf{a}_m^T \mathbf{X}\right] + b_m \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1^T \mathbb{E}[\mathbf{X}] + b_1 \\ \mathbf{a}_2^T \mathbb{E}[\mathbf{X}] + b_2 \\ \vdots \\ \mathbf{a}_m^T \mathbb{E}[\mathbf{X}] + b_m \end{pmatrix} = \mathbf{A}^T \mathbb{E}[\mathbf{X}] + \mathbf{b}$$

- Expected value operator can be interchanged with linear operations

Expected value of a binomial RV

- Binomial RVs count number of successes in n Bernoulli trials
- Let X_i $i = 1, \ldots n$ be n independent Bernouilli RVs
- Can write binomial X as $\Rightarrow X = \sum_{i=1}^{n} X_i$
- Expected value of binomial then $\Rightarrow \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np$
- Expected nr. successes = nr. trials $\times$ prob. individual success
  - Same interpretation that we observed for Poisson RVs
- Correlated Bernoulli trials $\Rightarrow X = \sum_{i=1}^{n} X_i$ but X_i are not independent
- Expected nr. successes is still $\mathbb{E}[X_i] = np$
  - Linearity of expectation does not require independence. Have not even defined independence for RVs yet

Independence

- Events E and F are independent if $\mathrm{P}[E \cap F] = \mathrm{P}[E]\,\mathrm{P}[F]$
- RVs X and Y are independent if events $X \leq x$ and $Y \leq y$ are independent for all x and y, i.e.

$$\mathrm{P}[X \leq x, Y \leq y] = \mathrm{P}[X \leq x]\,\mathrm{P}[Y \leq y]$$

- Obviously equivalent to $F_{XY}(x, y) = F_X(x)F_Y(y)$
- For discrete RVs equivalent to analogous relation between pmfs

$$p_{XY}(x, y) = F_X(x)F_Y(y)$$

- For continuous RVs the analogous is true for pdfs

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

Example: Sum of independent Poisson RVs

- Consider two Poisson RVs X and Y with parameters λ_x and λ_y
- Probability distribution of the sum RV $Z := X + Y$?

- $Z = n$ only if $X = k,\ Y = n - k$ for some $0 \leq k \leq n$ (independence, Poisson pmf definition, rearrange terms, binomial theorem)

$$
\begin{aligned}
p_Z(n) &= \sum_{k=0}^{n} \mathrm{P}\left[X = k, Y = n-k\right] \quad = \sum_{k=0}^{n} \mathrm{P}\left[X = k\right] \mathrm{P}\left[Y = n-k\right] \\
&= \sum_{k=0}^{n} e^{-\lambda_x} \frac{\lambda_x^k}{k!} e^{-\lambda_y} \frac{\lambda_y^{n-k}}{(n-k)!} \quad = \frac{e^{-(\lambda_x+\lambda_y)}}{n!} \sum_{k=0}^{n} \frac{n!}{(n-k)!k!} \lambda_x^k \lambda_y^{n-k} \\
&= \frac{e^{-(\lambda_x+\lambda_y)}}{n!} (\lambda_x + \lambda_y)^n
\end{aligned}
$$

- Z is Poisson with parameter $\lambda_z := \lambda_x + \lambda_y$

  $\Rightarrow$ Sum of independent Poissons is Poisson (parameters added)

Expected value of a product of independent RVs

Theorem

For independent RVs X and Y, and arbitrary functions $g(X)$ and $h(Y)$:

$$\mathbb{E}\left[g(X)h(Y)\right] = \mathbb{E}\left[g(X)\right]\mathbb{E}\left[h(Y)\right]$$

The expected value of the product is the product of the expected values

- As a particular case, when $g(X) = X$ and $h(Y) = Y$ we have

$$\mathbb{E}\left[XY\right] = \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$$

- Expectation and product can be interchanged if RVs are independent
- Different from interchange with linear operations (always possible)

Expected value of a product of independent RVs

Proof.

- For the case of X and Y continuos RVs. Use definition of independence to write

$$\mathbb{E}\left[g(X)h(Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y)\,dxdy$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)\,dxdy$$

- Integrand is product of a function of x and a function of y

$$\mathbb{E}\left[g(X)h(Y)\right] = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx \int_{-\infty}^{\infty} h(y)f_Y(y)\,dy$$
$$= \mathbb{E}\left[g(X)\right]\mathbb{E}\left[h(Y)\right]$$

□

Variance of a sum of independent RVs

- N Independent RVs $X_1, \ldots, X_N$
- Mean $\mathbb{E}[X_n] = \mu_n$ and Variance $\mathbb{E}\left[(X_n - \mu_n)^2\right] = \text{var}[X_n]$
- Variance of sum $X := \sum_{n=1}^{N} X_n$?
- Notice that mean of X is $\mathbb{E}[X] = \sum_{n=1}^{N} \mu_n$. Then

$$\text{var}[X] = \mathbb{E}\left[\left(\sum_{n=1}^{N} X_n - \sum_{n=1}^{N} \mu_n\right)^2\right] = \mathbb{E}\left[\left(\sum_{n=1}^{N} X_n - \mu_n\right)^2\right]$$

- Expand square and interchange summation and expectation

$$\text{var}[X] = \sum_{n=1}^{N} \sum_{m=1}^{N} \mathbb{E}\Big[(X_n - \mu_n)(X_m - \mu_m)\Big]$$

- Separate terms in sum. Use independence, definition of individual variances and $\mathbb{E}(X_n - \mu_n) = 0$

$$\text{var}[X] = \sum_{n=1, n\neq m}^{N} \sum_{m}^{N} \mathbb{E}\Big[(X_n - \mu_n)(X_m - \mu_m)\Big] + \sum_{n=1}^{N} \mathbb{E}\Big[(X_n - \mu_n)^2\Big]$$

$$= \sum_{n=1, n\neq m}^{N} \sum_{m}^{N} \mathbb{E}(X_n - \mu_n)\mathbb{E}(X_m - \mu_m) + \sum_{n=1}^{N} \mathbb{E}\Big[(X_n - \mu_n)^2\Big]$$

$$= \sum_{n=1}^{N} \text{var}[X_n]$$

- If variables are independent $\Rightarrow$ Variance of sum is sum of variances

Covariance

- The covariance of X and Y is (generalizes variance to pairs of RVs)

$$\text{cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

- If $\text{cov}(X, Y) = 0$ variables X and Y are said to be uncorrelated
- If X, Y independent then $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$ and $\text{cov}(X, Y) = 0$

  $\Rightarrow$ Independence implies uncorrelated RVs
- Opposite is not true, may have $\text{cov}(X, Y) = 0$ for dependent X, Y
  - E.g., X Uniform in $[-a, a]$ and $Y = X^2$
- But uncorrelation implies independence if X, Y are normal
- If $\text{cov}(X, Y) > 0$ then X and Y tend to move in the same direction

  $\Rightarrow$ Positive correlation
- If $\text{cov}(X, Y) < 0$ then X and Y tend to move in opposite directions

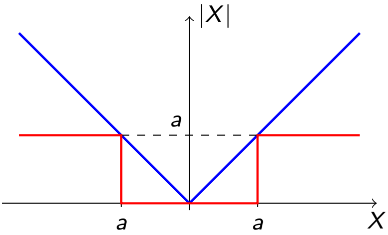
  $\Rightarrow$ Negative correlation

Markov and Chebyshev's Inequalities

Markov's inequality

- RV X with finite expected value $\mathbb{E}(X) < \infty$
- Markov's inequality states $\Rightarrow \mathrm{P}\left[|X| \geq a\right] \leq \dfrac{\mathbb{E}(|X|)}{a}$
- $\mathbb{I}\left\{|X| \geq a\right\} = 1$ when $X \geq a$ and 0 else. Then (figure to the right)

$$a\mathbb{I}\left\{|X| \geq a\right\} \leq |X|$$



- Expected value. Linearity of $\mathbb{E}\left[\cdot\right]$

$$a\mathbb{E}(\mathbb{I}\left\{|X| \geq a\right\}) \leq \mathbb{E}(|X|)$$

- Indicator function's expectation = Probability of event

$$a\mathrm{P}\left[|X| \geq a\right] \leq \mathbb{E}(|X|)$$

Chebyshev's inequality

- RV X with finite mean $\mathbb{E}(X) = \mu$ and variance $\mathbb{E}\left[(X-\mu)^2\right] = \sigma^2$
- Chebyshev's inequality $\Rightarrow \text{P}\left[|X-\mu| \geq k\right] \leq \dfrac{\sigma^2}{k^2}$
- Markov's inequality for the RV $Z = (X-\mu)^2$ and constant $a = k^2$

$$\text{P}\left[(X-\mu)^2 \geq k^2\right] = \text{P}\left[|Z| \geq k^2\right] \leq \frac{\mathbb{E}\left[|Z|\right]}{k^2} = \frac{\mathbb{E}\left[(X-\mu)^2\right]}{k^2}$$

- Notice that $(X-\mu)^2 \geq k^2$ if and only if $|X-\mu| \geq k$ thus

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

- Chebyshev's inequality follows from definition of variance

Comments & observations

- Markov and Chebyshev's inequalities hold for all RVs
- If absolute expected value is finite $\mathbb{E}[|X|] < \infty$

  $\Rightarrow$ RV's cdf decreases at least linearly (Markov's)
- If mean $\mathbb{E}(X)$ and variance $\mathbb{E}\left[(X-\mu)^2\right]$ are finite

  $\Rightarrow$ RV's cdf decreases at least quadratically (Chebyshev's)
- Most cdfs decrease exponentially (e.g. e^{-x^2} for normal)

  $\Rightarrow$ linear and quadratic bounds are loose but still useful

Limits

- Sequence of RVs $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$
- Distinguish between stochastic process $X_{\mathbb{N}}$ and realizations $x_{\mathbb{N}}$
- Say something about X_n for n large? $\Rightarrow$ Not clear, X_n is a RV
- Say something about x_n for n large? $\Rightarrow$ Certainly, look at $\lim_{n\to\infty} x_n$
- Say something about $\mathrm{P}[X_n]$ for n large? $\Rightarrow$ Yes, $\lim_{n\to\infty} \mathrm{P}[X_n]$
- Translate what we now about regular limits to definitions for RVs
- Can start from convergence of sequences: $\lim_{n\to\infty} x_n$
  - Sure and almost sure convergence
- Or from convergence of probabilities: $\lim_{n\to\infty} \mathrm{P}[X_n]$
  - Convergence in probability, mean square sense and distribution

- Denote sequence of variables $x_{\mathbb{N}} = x_1, x_2, \ldots, x_n, \ldots$
- Sequence $x_{\mathbb{N}}$ converges to the value x if given any $\epsilon > 0$

  $\Rightarrow$ There exists n_0 such that for all $n > n_0$, $|x_n - x| < \epsilon$
- Sequence x_n comes close to its limit $\Rightarrow |x_n - x| < \epsilon$
- And stays close to its limit $\Rightarrow$ for all $n > n_0$

- Stochastic process (sequence of RVs) $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$
- Realizations of $X_{\mathbb{N}}$ are sequences $x_{\mathbb{N}}$
- We say SP $X_{\mathbb{N}}$ converges surely to RV X if $\Rightarrow \lim_{n\to\infty} x_n = x$
- For all realizations $x_{\mathbb{N}}$ of $X_{\mathbb{N}}$

- Not really adequate. Even an event that happens with vanishingly small probability prevents sure convergence

Almost sure convergence

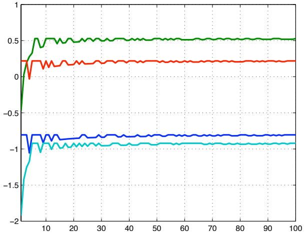
- RV X and stochastic process $X_{\mathbb{N}} = X_1, X_2, \ldots, X_n, \ldots$
- We say SP $X_{\mathbb{N}}$ converges almost surely to RV X if

$$\mathrm{P}\left[\lim_{n\to\infty} X_n = X\right] = 1$$

- Almost all sequences converge, except for a set of measure 0
- Almost sure convergence denoted as $\Rightarrow \lim_{n\to\infty} X_n = X \quad \text{a.s.}$
- Limit X is a random variable

Example

- $X_0 \sim \mathcal{N}(0,1)$ (normal, mean 0, variance 1)
- Z_n Bernoulli parameter p
- Define $\Rightarrow X_n = X_0 - \dfrac{Z_n}{n}$
- $Z_n/n \to 0$, then $\lim_{n\to\infty} X_n = X_0$ a.s.



Convergence in probability

- We say SP $X_{\mathbb{N}}$ converges in probability to RV X if for any $\epsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}\left[|X_n - X| < \epsilon\right] = 1$$

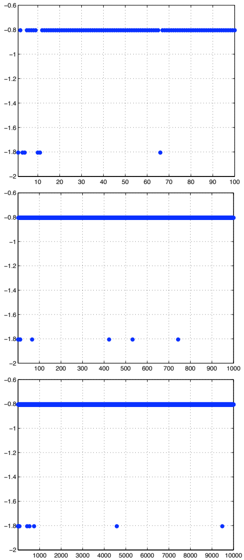
- Probability of distance $|X_n - X|$ becoming smaller than ϵ tends to 1
- Statement is about probabilities, not about processes
- The probability converges
- Realizations $x_{\mathbb{N}}$ of $X_{\mathbb{N}}$ might or might not converge
- Limit and probability interchanged with respect to a.s. convergence
- a.s. convergence implies convergence in probability
  - If $\lim_{n\to\infty} X_n = X$ then for any $\epsilon > 0$ there is n_0 such that $|X_n - X| < \epsilon$ for all $n \geq n_0$
  - This is true for all almost all sequences then $\mathrm{P}\left[|X_n - X| < \epsilon\right] \to 1$

Example

- $X_0 \sim \mathcal{N}(0,1)$ (normal, mean 0, variance 1)
- Z_n Bernoulli parameter $1/n$
- Define $\Rightarrow X_n = X_0 - Z_n$
- X_n converges in probability to X_0 because

$$\begin{aligned} \mathrm{P}\left[|X_n - X_0| < \epsilon\right] &= \mathrm{P}\left[|Z_n| < \epsilon\right] \\ &= 1 - \mathrm{P}\left[Z_n = 1\right] \\ &= 1 - \frac{1}{n} \to 1 \end{aligned}$$

- Plot of path x_n up to $n = 10^2$, $n = 10^3$, $n = 10^4$
- $Z_n = 1$ becomes ever rarer but still happens



Difference between a.s. and p

- Almost sure convergence implies that almost all sequences converge
- Convergence in probability does not imply convergence of sequences

- Latter example: $X_n = X_0 - Z_n$, Z_n is Bernoulli with parameter $1/n$
- As we've seen it converges in probability

$$\mathrm{P}\left[|X_n - X_0| < \epsilon\right] = 1 - \frac{1}{n} \to 1$$

- But for almost all sequences, the $\lim_{n\to\infty} X_n$ does not exist

- Almost sure convergence $\Rightarrow$ disturbances stop happening
- Convergence in prob. $\Rightarrow$ disturbances happen with vanishing freq.
- Difference not irrelevant.
  - Interpret Y_n as rate of change in savings
  - with a.s. convergence risk is eliminated
  - with convergence in probability risk decreases but does not disappear

Mean square convergence

- We say SP $X_{\mathbb{N}}$ converges in mean square to RV X if

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^2\right] = 0$$

- Sometimes (very) easy to check
- Convergence in mean square implies convergence in probability
- From Markov's inequality

$$\text{P}\left[|X_n - X| \geq \epsilon\right] = \text{P}\left[|X_n - X|^2 \geq \epsilon^2\right] \leq \frac{\mathbb{E}\left[|X_n - X|^2\right]}{\epsilon^2}$$

- If $X_n \to X$ in mean square sense, $\mathbb{E}\left[|X_n - X|^2\right]/\epsilon^2 \to 0$ for all ϵ
- Almost sure and mean square $\nRightarrow$ neither implies the other

Convergence in distribution

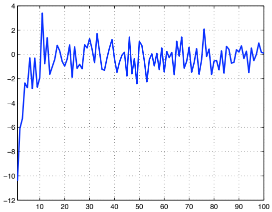
- Stochastic process $X_{\mathbb{N}}$. Cdf of X_n is $F_n(x)$
- The SP converges in distribution to RV X with distribution $F_X(x)$ if

$$\lim_{n\to\infty} F_n(x) = F_X(x)$$

- For all x at which $F_X(x)$ is continuous
- Again, no claim about individual sequences, just the cdf of X_n
- Weakest form of convergence covered,
- Implied by almost sure, in probability, and mean square convergence

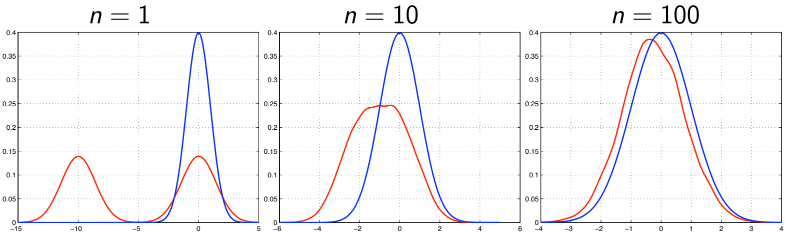
Example

- $Y_n \sim \mathcal{N}(0, 1)$
- Z_n Bernoulli parameter p
- Define $\Rightarrow X_n = Y_n - 10Z_n/n$
- $Z_n/n \to 0$, then $\lim_{n\to\infty} F_n(x) = \mathcal{N}(0, 1)$



Convergence in distribution (continued)

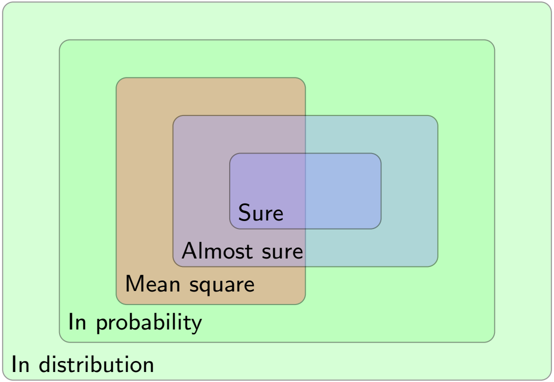
- Individual sequences x_n do not converge in any sense
  $\Rightarrow$ It is the distribution that converges



- As the effect of Z_n/n vanishes pdf of X_n converges to pdf of Y_n
  - Standard normal $\mathcal{N}(0, 1)$

Implications

- Sure $\Rightarrow$ almost sure $\Rightarrow$ in probability $\Rightarrow$ in distribution
- Mean square $\Rightarrow$ in probability $\Rightarrow$ in distribution
- In probability $\Rightarrow$ in distribution



Limit theorems

Sum of independent identically distributed RVs

- Independent identically distributed (i.i.d.) RVs $X_1, X_2, \ldots, X_n, \ldots$
- Mean $\mathbb{E}[X_n] = \mu$ and variance $\mathbb{E}\left[(X_n - \mu)^2\right] = \sigma^2$ for all n
- What happens with sum $S_N := \sum_{n=1}^{N} X_n$ as N grows?

- Expected value of sum is $\mathbb{E}[S_N] = N\mu \;\Rightarrow$ Diverges if $\mu \neq 0$
- Variance is $\mathbb{E}\left[(S_N - N\mu)^2\right] = N\sigma$

  $\Rightarrow$ Diverges if $\sigma \neq 0$ (alwyas true unless X_n is a constant)

- One interesting normalization $\Rightarrow \bar{X}_N := (1/N)\sum_{n=1}^{N} x_n$
- Now $\mathbb{E}[Z_N] = \mu$ and $\text{var}[Z_N] = \sigma^2/N$
- Law of large numbers (weak and strong)

- Another interesting normalization $\Rightarrow Z_N := \dfrac{\sum_{n=1}^{N} x_n - N\mu}{\sigma\sqrt{N}}$
- Now $\mathbb{E}[Z_N] = 0$ and $\text{var}[Z_N] = 1$ for all values of N
- Central limit theorem

Weak law of large numbers

- i.i.d. sequence or RVs $X_1, X_2, \ldots, X_n, \ldots$ with mean $\mu = \mathbb{E}[X_n]$
- Define sample average $\bar{X}_N := (1/N)\sum_{n=1}^{N} x_n$

- Weak law of large numbers
- Sample average $\bar{X}_N$ converges in probability to $\mu = \mathbb{E}[X_n]$

$$\lim_{N\to\infty} \mathrm{P}\left[|\bar{X}_N - \mu| > \epsilon\right] = 1, \quad \text{for all } \epsilon > 0$$

- Strong law of large numbers
- Sample average $\bar{X}_N$ converges almost surely to $\mu = \mathbb{E}[X_n]$

$$\mathrm{P}\left[\lim_{N\to\infty} \bar{X}_N = \mu\right] = 1$$

- Strong law implies weak law. Can forget weak law if so wished

Proof of weak law of large numbers

- Weak law of large numbers is very simple to prove

Proof.

- Variance of $\bar{X}_n$ vanishes for N large

$$\text{var}\left[\bar{X}_N\right] = \frac{1}{N^2}\sum_{n=1}^{n}\text{var}\left[X_n\right] = \frac{\sigma^2}{N} \to 0$$

- But, what is the variance of $\bar{X}_N$?

$$0 \leftarrow \frac{\sigma^2}{N} = \text{var}\left[\bar{X}_N\right] = \mathbb{E}\left[(\bar{X}_n - \mu)^2\right]$$

- Then, $|\bar{X}_N - \mu|$ converges in mean square sense

  $\Rightarrow$ Which implies convergence in probability □

- Strong law is a little more challenging

Central limit theorem (CLT)

Theorem

- *i.i.d. sequence of RVs* $X_1, X_2, \ldots, X_n, \ldots$
- *Mean* $\mathbb{E}\left[X_n\right] = \mu$ *and variance* $\mathbb{E}\left[(X_n - \mu)^2\right] = \sigma^2$ *for all* n
- *Then* $\Rightarrow \lim_{N\to\infty} \mathrm{P}\left[\frac{\sum_{n=1}^{N} x_n - N\mu}{\sigma\sqrt{N}} \leq x\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}\, du$

- Former statement implies that for N sufficiently large

$$Z_N := \frac{\sum_{n=1}^{N} x_n - N\mu}{\sigma\sqrt{N}} \sim \mathcal{N}(0,1)$$

- $\sim$ means "distributed like"
- Z_N converges in distribution to a standard normal RV

CLT (continued)

- Equivalently can say $\Rightarrow \sum_{n=1}^{N} x_n \sim \mathcal{N}(N\mu, N\sigma^2)$
- Sum of large number of i.i.d. RVs has a normal distribution
  - Cannot take a meaningful limit here.
  - But intuitively, this is what the CLT states

Example

- Binomial RV X with parameters (n, p)
- Write as $X = \sum_{i=1}^{n} X_i$ with X_i Bernoulli with parameter p
- Mean $\mathbb{E}[X_i] = p$ and variance $\text{var}[X_i] = p(1-p)$
- For sufficiently large $n \Rightarrow X \sim \mathcal{N}(n\mu, np(1-p))$