Probability review

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

September 9, 2011

Definitions

Sigma algebras and probability spaces

Conditional probability, independence, total probability, Bayes's rule

Random variables

Discrete random variables

- Commonly used discrete random variables

Continuous random variables

- Commonly used continuous random variables

Expected values

- Discrete random variables
- Continuous random variables
- Functions of random variables

Joint probability distributions

Joint expectations

- Independence

Probability

- An event is a thing that happens
- A random event is one that is not certain
- The probability of an event measures how likely it is to occur

Example

- I've written a student's name in a piece of paper. Who is she/he?
- Event(s): Student x's name is written in the paper
- Probability(ies): $P(x)$ how likely is x's name to be the one written

- Probability is a measurement tool

Sigma-Algebra

- Given a space or universe S
  - E.g., all students in the class $S = \{x_1, x_2, \ldots, x_N\}$ (x_n denote names)
- An event E is a subset of S
  - E.g. $\{x_1\}$, student with name x_1,
  - Or in general $\{x_n\}$, student with name x_n
  - But also $\{x_1, x_4\}$, students with names x_1 and x_4
- A sigma-Algebra $\mathcal{F}$ is a collection of events $E \subseteq S$ such that
  - Not empty: $\mathcal{F} \neq \emptyset$
  - Closed under complement: If $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$
  - Closed under countable unions: If $E_i \in \mathcal{F}$ $\cup_{i=1}^{\infty} E_i \in \mathcal{F}$
- Note that $\mathcal{F}$ is a set of sets

Examples of Sigma-Algebras

Example

- No student and all students, i.e., $\mathcal{F}_0 := \{\emptyset, S\}$

Example

- Empty set, women, men, all students, i.e., $\mathcal{F}_1 := \{\emptyset, \text{Women}, \text{Men}, S\}$

Example

- $\mathcal{F}$ including the empty set plus
- All events (sets) with one student $\{x_1\}, \ldots, \{x_N\}$ plus
- All events with two students $\{x_1, x_2\}$, $\{x_1, x_3\}$, ..., $\{x_1, x_N\}$,
  $\{x_2, x_3\}$, ..., $\{x_2, x_N\}$,
  ...
  $\{x_{N-1}, x_N\}$ plus
- All events with three students, four, ..., N students.

Axioms of probability

- Define a function $P(E)$ from a sigma-Algebra $\mathcal{F}$ to the real numbers
- $P(E)$ is a probability if
  - $\Rightarrow$ Probability range $\Rightarrow 0 \leq P(E) \leq 1$
  - $\Rightarrow$ Probability of universe $\Rightarrow P(S) = 1$
  - $\Rightarrow$ Additivity $\Rightarrow$ Given sequence of disjoint events $E_1, E_2, \ldots$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P\left(E_i\right)$$

  - $\Rightarrow$ Probability of union is the sum of individual probabilities
- In additivity property number of events is possibly infinite
- Disjoint events means $E_i \cap E_j = \emptyset$

Probability example

- Sigma-algebra with all combinations of students
- Names are equiprobable $\Rightarrow P(x_n) = 1/N$ for all n.

  $\Rightarrow$ Is this function a probability? Is there enough information given?

- Sets with two students (for $n \neq m$):

$$P(\{x_n, x_m\}) = P(\{x_n\}) + P(\{x_m\}) = 2/N$$

  $\Rightarrow$ Is this function a probability? Is there enough information given?

- Have to specify probability for all elements of the sigma-algebra

  $\Rightarrow$ Sets with 3 students $\Rightarrow 3/N$. Sets with 4 students $\Rightarrow 4/N$ ...

  $\Rightarrow$ For universe $S \Rightarrow P(S) = P\left(\bigcup_{n=1}^{N}\{x_n\}\right) = \sum_{i=1}^{\infty} P(x_n) = 1 = N/N$

- Is this function a probability? $\Rightarrow$ Verify properties (range, universe, additivity)

Definitions

- Sigma algebras and probability spaces
- Conditional probability, independence, total probability, Bayes's rule
- Random variables
- Discrete random variables
  - Commonly used discrete random variables
- Continuous random variables
  - Commonly used continuous random variables
- Expected values
  - Discrete random variables
  - Continuous random variables
  - Functions of random variables
- Joint probability distributions
- Joint expectations
  - Independence

Conditional probability

- Partial information about the event (E.g. Name is male)
- The event E belongs to a set F
- Define the conditional probability of E given F as

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

- Renormalize probabilities to the set F
- Need to have $P(F) > 0$

Conditional probability example

- The name I wrote is male. What is the probability of name x_n?
- Assume male names are $F = \{x_1, \ldots, x_M\}$
- Probability of F is $P(F) = M/N$

- If name is male, $x_n \in F$ and we have for event $E = \{x_n\}$

$$P(E \cap F) = P(\{x_n\}) = 1/N$$

- Conditional probability is as you would expect

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)} = \frac{1/N}{M/N} = \frac{1}{M}$$

- If name is female $x_n \notin F$, then $P(E \cap F) = P(\emptyset) = 0$
- As you would expect, then $P(E \mid F) = 0$

Independence

- Events E and F are said independent if $P(E \cap F) = P(E)P(F)$
- According to definition of conditional probability

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E)P(F)}{P(F)} = P(E)$$

- Knowing F does not alter our perception of E
- F has no information about E
- The symmetric is also true $P(F \mid E) = P(F)$
- Events that are not independent are dependent

Independence example

- Wrote one name, asked a friend to write another
- Space S is sets of all pairs of names $[x_n(1), x_n(2)]$
- Sigma-algebra is cartesian product $\mathcal{F} \times \mathcal{F}$
- Pair of names chosen without coordination

$$P\big(\{(x_1, x_2)\}\big) = P\big(\{x_1\}\big)P\big(\{x_2\}\big) = \frac{1}{N^2}$$

- Dependent events: I wrote one name, then another name

Total probability

- Consider event E and events F and F^c
- F and F^c are a partition of the space S ($F \cup F^c = S$, $F \cap F^c = \emptyset$)
- Because $F \cup F^c = S$ cover space S can write the set E as

$$E = E \cap S = E \cap (F \cup F^c) = (E \cap F) \cup (E \cap F^c)$$

- Because $F \cap F^c = \emptyset$ are disjoint, so is $(E \cap F) \cap (E \cap F^c) = \emptyset$. Thus

$$\mathrm{P}[E] = \mathrm{P}[(E \cap F) \cup (E \cap F^c)] = \mathrm{P}[E \cap F] + \mathrm{P}[E \cap F^c]$$

- Use definition of conditional probability

$$\mathrm{P}[E] = \mathrm{P}\left[E \,\middle|\, F\right] \mathrm{P}[F] + \mathrm{P}\left[E \,\middle|\, F^c\right] \mathrm{P}[F^c]$$

- Translate conditional information, $\mathrm{P}\left[E \,\middle|\, F\right]$ and $\mathrm{P}\left[E \,\middle|\, F^c\right]$
  - $\Rightarrow$ Into unconditional information $\mathrm{P}[E]$

Total probability - continued

- In general, consider (possibly infinite) partition F_i, $i = 1, 2, \ldots$ of S
- Sets F_i are disjoint $\Rightarrow F_i \cap F_j = \emptyset$ for $i \neq j$
- Sets F_i cover the space $\Rightarrow \cup_{i=1}^{\infty} F_i = S$

- As before, because $\cup_{i=1}^{\infty} F_i = S$ cover space S can write the set E as

$$E = E \cap S = E \cap \left(\bigcup_{i=1}^{\infty} F_i \right) = \bigcup_{i=1}^{\infty} E \cap F_i$$

- Because $F_i \cap F_j = \emptyset$ are disjoint, so is $(E \cap F_i) \cap (E \cap F_j) = \emptyset$. Thus

$$\mathrm{P}[E] = \mathrm{P}\left[\bigcup_{i=1}^{\infty} E \cap F_i \right] = \sum_{i=1}^{\infty} \mathrm{P}[E \cap F_i] = \sum_{i=1}^{\infty} \mathrm{P}\left[E \mid F_i \right] \mathrm{P}[F_i]$$

Total probability example

- In this class seniors get an A with probability 0.9
- Juniors get an A with probability 0.8
- For a exchange student, we estimate its standing as being senior with prob. 0.7 and junior with prob. 0.3
- What is the probability of the exchange student scoring an A?

- Let $A =$ "exchange student gets an A," S denote senior standing and J junior standing
- Use total probability

$$\mathrm{P}\left[A\right] = \mathrm{P}\left[A \,\middle|\, S\right] \mathrm{P}\left[S\right] + \mathrm{P}\left[A \,\middle|\, J\right] \mathrm{P}\left[J\right]$$

- Or in numbers

$$\mathrm{P}\left[A\right] = 0.9 \times 0.7 + 0.8 \times 0.3 = 0.87$$

Bayes's Rule

- From the definition of conditional probability

$$P(E \mid F)P(F) = P(E \cap F)$$

- Likewise, for F conditioned on E, we have

$$P(F \mid E)P(E) = P(F \cap E)$$

- Quantities above are equal, then

$$P(E \mid F) = \frac{P(F \mid E)P(E)}{P(F)}$$

- Bayes's rule allows time reversion. If F (future) comes after E (past),
  - $\Rightarrow P(E \mid F)$, probability of past (E) having seen the future (F)
  - $\Rightarrow P(F \mid E)$, probability of future (F) having seen past (E)
- Models often describe future $\mid$ past. Interest is often in past $\mid$ future

Random variables

Sigma algebras and probability spaces

Conditional probability, independence, total probability, Bayes's rule

Random variables

Discrete random variables

- Commonly used discrete random variables

Continuous random variables

- Commonly used continuous random variables

Expected values

- Discrete random variables
- Continuous random variables
- Functions of random variables

Joint probability distributions

Joint expectations

- Independence

Random variables (RV) definition

- A RV X is a function that assigns a number to a random event
- Think of RVs as measurements.
- Event is something that happens, RV is an associated measurement
- Probabilities of RVs inferred from probabilities of underlying events

Example

- Throw a ball inside a $1m \times 1m$ square. Interested in ball position
- Random event is the place where the ball falls
- Random variables are x and y position coordinates

Example 1

- Throw coin for head (H) or tail (T). Coin is fair $\mathrm{P}[H] = 1/2$, $\mathrm{P}[T] = 1/2$. Pay \$1 for H, charge \$1 for T. Earnings?
- Events are H and T
- To measure earnings define RV X with values

$$X(H) = 1, \qquad X(T) = -1$$

- Probabilities of the RV are

$$\begin{aligned} \mathrm{P}[X = 1] \quad &= \mathrm{P}[H] = 1/2, \\ \mathrm{P}[X = -1] &= \mathrm{P}[T] = 1/2 \end{aligned}$$

- We also have $\mathrm{P}[X = a] = 0$ for all other $a \neq 1$, $a \neq -1$

Example 2

- Throw 2 coins. Pay \$1 for each H, charge \$1 for each T.
- Events are HH and HT, TH, TT
- To measure earnings define RV Y with values

$$Y(HH) = 2, \quad Y(HT) = 0, \quad Y(TH) = 0, \quad Y(TT) = -2$$

- Probabilities are

$$\begin{aligned} \mathrm{P}[X = 2] &= \mathrm{P}[HH] &&= 1/4, \\ \mathrm{P}[X = 0] &= \mathrm{P}[HT] + \mathrm{P}[TH] &&= 1/2, \\ \mathrm{P}[X = -2] &= \mathrm{P}[HT] &&= 1/4, \end{aligned}$$

About Examples 1 & 2

- RVs are easier to manipulate than events
- Let $E_1 \in \{H, T\}$ be outcome of coin 1 and $E_2 \in \{H, T\}$ of coin 2
- Can relate X and Y as

$$Y(E_1, E_2) = X(E_1) + X(E_2)$$

- Throw N coins. Earnings?
- Enumeration becomes cumbersome
- Let $E_n \in \{H, T\}$ be outcome of n-th coin and define

$$Y(E_1, E_2, \ldots, E_n) = \sum_{n=1}^{N} X(E_n)$$

Example 3

- Throw a coin until landing heads for the first time. $P(H) = p$
- Number of throws until the first head?
- Events are H, TH, TTH, $TTTH$, ...
  - We stop throwing coins at first head (thus THT not a possible event)
- Let N be RV with number of throws.
- $N = n$ if we land T in the first $n-1$ throws and and H in the n-th

$$\begin{aligned}
\mathrm{P}[N=1] &= \mathrm{P}[H] && = p \\
\mathrm{P}[N=2] &= \mathrm{P}[TH] && = (1-p)p \\
&\vdots \\
\mathrm{P}[X=n] &= \mathrm{P}[TT\ldots TH] && = (1-p)^{n-1}p
\end{aligned}$$

Example 3 - continued

- It should be $\sum_{n=1}^{\infty} \mathrm{P}\left[N=n\right] = 1$
- This is true because $\sum_{n=1}^{\infty}(1-p)^{n-1}$ is a geometric sum. Then

$$\sum_{n=1}^{\infty}(1-p)^{n-1} = 1 + (1-p) + (1-p)^2 + \ldots = \frac{1}{1-p}$$

- Using this for the sum of probabilities

$$\sum_{n=1}^{\infty} \mathrm{P}\left[N=n\right] = p\sum_{n=1}^{\infty}(1-p)^{n} = p\,\frac{1}{1-(1-p)} = 1.$$

Indicator function

- The indicator function is a random variable
- Let E be an event. Let e be the outcome of a random event

$$\mathbb{I}\{E\} = 1 \quad \text{if } e \in E$$
$$\mathbb{I}\{E\} = 0 \quad \text{if } e \notin E$$

- It indicates that outcome e belongs to set E, by taking value 1

Example

- Number of throws N until first H. Interested on N exceeding N_0
- Event is $\{N : N > N_0\}$. Possible outcomes are $N = 1, 2, \ldots$
- Denote indicator function as $\Rightarrow \mathbb{I}_{N_0} = \mathbb{I}\{N : N > N_0\}$
- The probability $\mathrm{P}[\mathbb{I}_{N_0} = 1] = \mathrm{P}[N > N_0] = (1-p)^{N_0}$

  $\Rightarrow$ For N to exceed N_0 need N_0 consecutive tails

  $\Rightarrow$ Doesn't matter what happens afterwards

Discrete random variables

Sigma algebras and probability spaces

Conditional probability, independence, total probability, Bayes's rule

Random variables

Discrete random variables

- Commonly used discrete random variables

Continuous random variables

- Commonly used continuous random variables

Expected values

- Discrete random variables
- Continuous random variables
- Functions of random variables

Joint probability distributions

Joint expectations

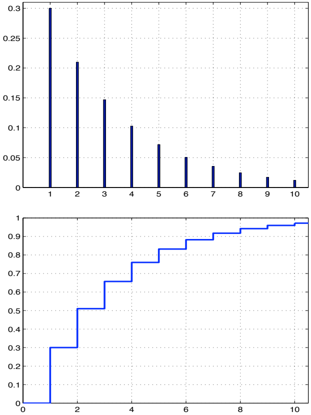
- Independence

- A discrete RV takes on, at most, a countable number of values
- Probability mass function (pmf) $p_X(x) = \mathrm{P}[X = x]$
  - If the RV is clear from context we just write $p_X(x) = p(x)$

- If X take values in $\{x_1, x_2, \ldots\}$ pmf satisfies
  - $p(x_i) > 0$ for $i = 1, 2, \ldots$
  - $p(x) = 0$ for all other $x \neq x_i$
  - $\sum_{i=1}^{\infty} p(x_i) = 1$
- Pmf for "throw to first head" (p=0.3)
- Cumulative distribution function (cdf) is

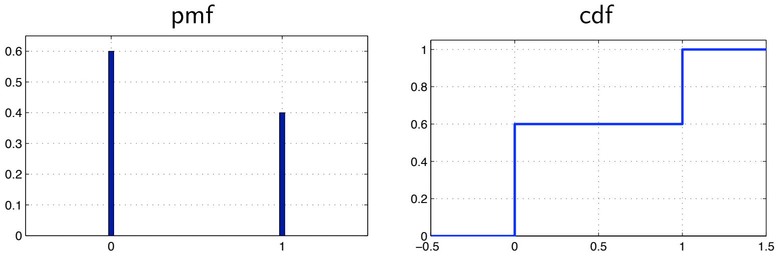
$$F_X(x) = \mathrm{P}[X \leq x] = \sum_{i: x_i \leq x} p(x_i)$$

- Staircase function with jumps at each x_i
- Cdf for "throw to first head" (p=0.3)



Bernoulli

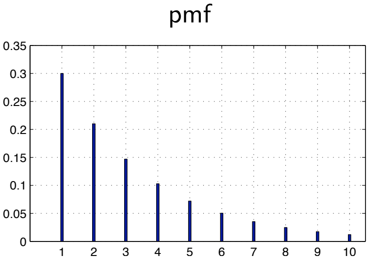
- An experiment/bet can succeed with probability p or fail with probability $(1-p)$ (e.g., coin throws, any indication of an event)
- Bernoulli X can be 0 or 1. Pmf values $\Rightarrow p(1) = p$
  $\Rightarrow p(0) = q = 1 - p$
- For the cdf we have $\Rightarrow F(x) = 0 \ \text{ for } \ x < 0$
  $\Rightarrow F(x) = q \ \text{ for } \ 0 \leq x < 1$
  $\Rightarrow F(x) = 1 \ \text{ for } \ 1 < x$

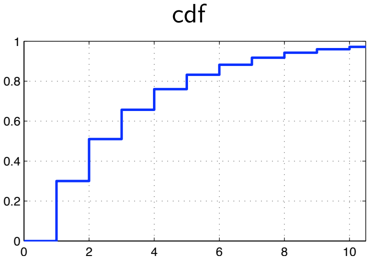


Geometric

- Count number of Bernoulli trials needed to register first success

- Trials succeed with probability p
- Number of trials X until success is geometric with parameter p
- Pmf is $\Rightarrow p(i) = p(1-p)^{i-1}$
  - $i-1$ failures plus one success. Throws are independent
- Cdf is $\Rightarrow F(i) = 1-(1-p)^{i-1}$
  - reaches i only if first $i-1$ trials fail; or just sum the geometric series

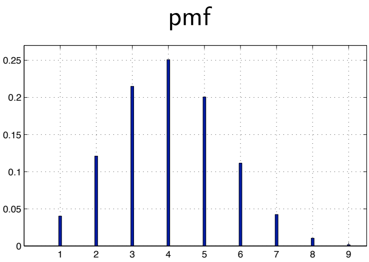


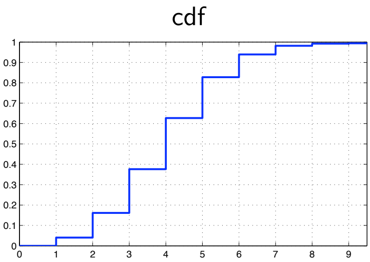


- Count number of successes X in n Bernoulli trials
- n trials. Probability of success p. Probability of failure $q = 1 - p$
- Then, binomial X with parameters (n, p) has pmf

$$p(i) = \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i}$$

- $X = i$ if there are i successes (p^i) and $n - i$ failures ($(1-p)^{n-i}$).
- There are $\binom{n}{i}$ ways of drawing i successes and $n - i$ failures





Binomial continued

- Let Y_i for $i = 1, \ldots n$ be Bernoulli RVs with parameter p
  $\Rightarrow$ Y_i associated with independent events
- Can write binomial X with parameters (n, p) as $\Rightarrow X = \sum_{i=1}^{n} Y_i$
- Consider binomials Y and Z with parameters (n_Y, p) and (n_Z, p)
- Probability distribution of $X = Y + Z$?
- Write $Y = \sum_{i=1}^{n_Y} Y_i$ and $Z = \sum_{i=1}^{n_Z} Z_i$ with Y_i and Z_i Bernoulli with parameter p. Write X as

$$X = \sum_{i=1}^{n_Y} Y_i + \sum_{i=1}^{n_Z} Z_i$$

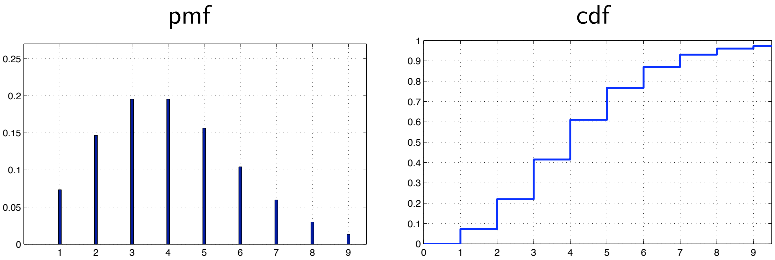
- Then X is binomial with parameter $(n_Y + n_Z, p)$

- Approximate a Binomial variable for large n

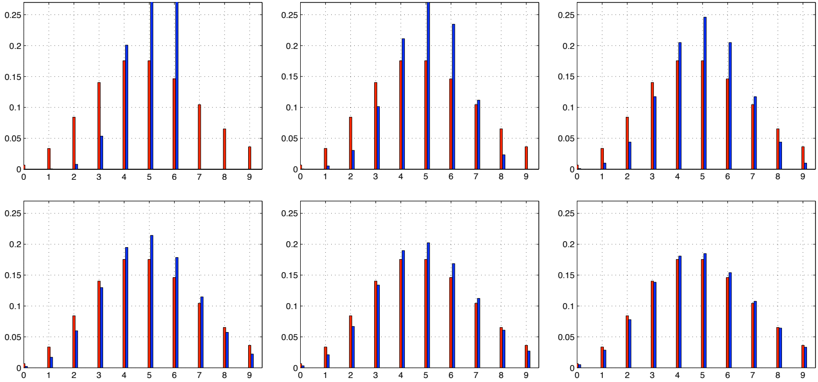
$$p(i) = e^{-\lambda}\frac{\lambda^i}{i!}$$

- Is this a properly defined pmf? Yes
- Taylor's expansion of $e^x = 1 + x + x^2/2 + \ldots + x^i/i! + \ldots$. Then

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{-\lambda}e^{\lambda} = 1$$



- X is binomial with parameters (n, p)
- Let $n \to \infty$ while maintaining a constant product $np = \lambda$
  - If we just let $n \to \infty$ number of successes diverges. Boring.
- Compare with Poisson distribution with parameter λ
  - $\lambda = 5$ $n = 6, 8, 15, 20, 50$



Poisson and binomial - continued

- This is, in fact, the motivation for the definition of a Poisson RV
- Substituting $p = \lambda/n$ in the pmf of a binomial RV

$$p_n(i) = \frac{n!}{(n-i)!i!}\left(\frac{\lambda}{n}\right)^i\left(1-\frac{\lambda}{n}\right)^{n-i}$$
$$= \frac{n(n-1)\dots(n-i+1)}{n^i}\,\frac{\lambda^i}{i!}\,\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}$$

- Factorials' defs., $(1-\lambda/n)^{n-i} = (1-\lambda/n)^n/(1-\lambda/n)^i)$, reorder terms
- By definition red term is $\lim_{n\to\infty}(1-\lambda/n)^n = e^{-\lambda}$
- Black and blue terms converge to 1. From both observations

$$\lim_{n\to\infty} p_n(i) = 1\,\frac{\lambda^i}{i!}\,\frac{e^{-\lambda}}{1} = e^{-\lambda}\frac{\lambda^i}{i!}$$

- Limit is the pmf of a Poisson RV

Poisson and binomial - continued

- Binomial distribution is justified by counting successes
- The Poisson is an approximation for large number of trials n
- Poisson distribution is more tractable
- Sometimes called "law of rare events"
  - Individual events (successes) happen with small probability $p = \lambda/n$
  - The aggregate event, though, (number of successes) need not be rare
- Notice that all four RVs are related to coin tosses.

Continuous random variables

Sigma algebras and probability spaces

Conditional probability, independence, total probability, Bayes's rule

Random variables

Discrete random variables

- Commonly used discrete random variables

Continuous random variables

- Commonly used continuous random variables

Expected values

- Discrete random variables
- Continuous random variables
- Functions of random variables

Joint probability distributions

Joint expectations

- Independence

Continuous RVs, probability density function

- Possible values for continuous RV X form a dense subset $\mathcal{X} \in \mathbb{R}$
- Uncountable infinite number of possible values

  $\Rightarrow$ May have $\text{P}\left[X = x\right] = 0$ for all $x \in \mathcal{X}$ (most certainly will)

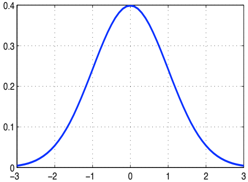
- The probability density function (pdf) is a function such that for any subset $\mathcal{X} \in \mathbb{R}$ (Normal pdf to the right)

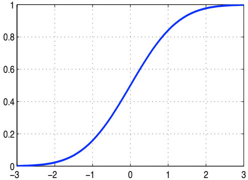
$$\text{P}\left[X \in \mathcal{X}\right] = \int_{\mathcal{X}} f_X(x)$$

- Cdf can be defined as before and related to the pdf (Normal cdf to the right)

$$F_X(x) = \Pr\left[X \leq x\right] = \int_{-\infty}^{x} f_X(u)\, du$$

- $\text{P}\left[X \leq \infty\right] = F_X(\infty) = \lim_{x\to\infty} F_X(x) = 1$





More on cdfs and pdfs

- When the set $\mathcal{X} = [a, b]$ is an interval of the real line

$$\mathrm{P}\left[X \in [a, b]\right] = \mathrm{P}\left[X \leq b\right] - \mathrm{P}\left[X \leq a\right] = F_X(b) - F_X(a)$$

- Or in terms of the pdf can be written as

$$\mathrm{P}\left[X \in [a, b]\right] = \int_a^b f_X(x)\, dx$$

- For small interval $[x_0, x_0 + \delta x]$, in particular

$$\mathrm{P}\left[X \in [x_0, x + \delta x]\right] = \int_{x_0}^{x+\delta x} f_X(x)\, dx \approx f_X(x_0)\delta x$$

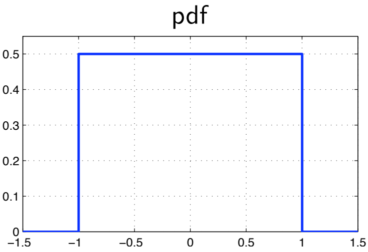
- Probability is the "area under the pdf" (thus "density")

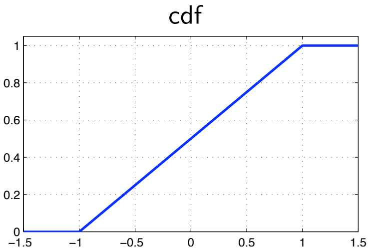
- Another relationship between pdf and cdf is $\Rightarrow \dfrac{\partial F_X(x)}{\partial x} = f_X(x)$
- From fundamental theorem of calculus ("derivative inverse of integral")

- Model problems with equal probability of landing on an interval $[a, b]$
- Pdf is $f(x) = 0$ outside the interval $[a, b]$ and

$$f(x) = \frac{1}{b-a}, \text{ for } a \leq x \leq b$$

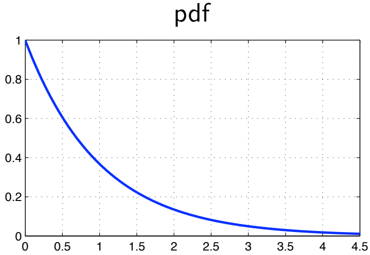
- Cdf is $F(x) = (x-a)/(b-a)$ in the interval $[a, b]$ (0 before, 1 after)
- Prob. of interval $[\alpha, \beta] \subseteq [a, b]$ is $\int_\alpha^\beta f(x) = (\beta - \alpha)/(b-a)$
  - $\Rightarrow$ Depends on interval's width $\beta - \alpha$ only, Not on its position





- Model memoryless times (more later)
- Pdf is $f(x) = 0$ for $x < 0$ and $f(x) = \lambda e^{-\lambda x}$ for $0 \leq x$
- CDF obtained by integrating pdf

$$F(x) = \int_{-\infty}^{x} f(u)\, du = \int_{0}^{x} \lambda e^{-\lambda u}\, du = -e^{-\lambda u}\Big|_{0}^{x} = 1 - e^{-\lambda x}$$

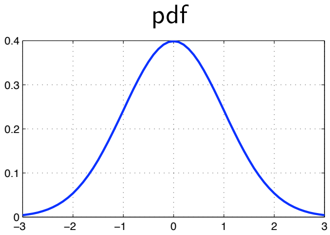


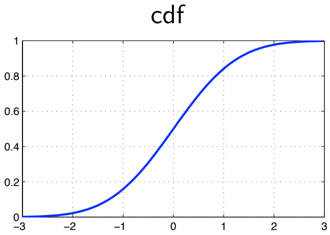
Normal / Gaussian

- Appears in phenomena where randomness arises from a large number of small random effects. Pdf is

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

- μ is the mean of the Normal RV. Shifts pdf to right ($\mu > 0$) or left
- σ^2 is the variance, σ the standard deviation. Controls width of pdf
  - 0.68 prob. between $\mu \pm \sigma$, 0.997 prob. in $\mu \pm 3\sigma$
- The cdf $F(x)$ cannot be expressed in terms of elementary functions





Expected values

Sigma algebras and probability spaces

Conditional probability, independence, total probability, Bayes's rule

Random variables

Discrete random variables
- Commonly used discrete random variables

Continuous random variables
- Commonly used continuous random variables

Expected values
- Discrete random variables
- Continuous random variables
- Functions of random variables

Joint probability distributions

Joint expectations
- Independence

Expected values

- We are asked to condense information about a RV in a single value
- What should this value be?
- If we are allowed a description with a few values
- What should they be?

- Expected values are convenient answers to these questions

- Beware: Expectations are condensed descriptions
- They necessarily overlook some aspects of the random phenomenon

Definition for discrete RVs

- RV X taking on values x_i, $i = 1, 2, \ldots$ with pmf $p(x)$
- The expected value of the RV X is

$$\mathbb{E}[X] := \sum_{i=1}^{\infty} x_i p(x_i) = \sum_{x:p(x)>0} x p(x)$$

- Weighted average of possible values x_i
- Common average if RV takes values x_i, $i = 1, \ldots, N$ equiprobably

$$\mathbb{E}[X] = \sum_{i=1}^{N} x_i p(x_i) = \sum_{i=1}^{N} x_i \frac{1}{N} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Expected value of Bernoulli and geometric RVs

- For a Bernoulli RV $p(1) = p$, $p(0) = 1 - p$ $p(x) = 0$ elsewhere
- Expected value $\Rightarrow \mathbb{E}[X] = (1)p + (0)q = p$

- For a geometric RV $p(x) = p(1-p)^{x-1} = pq^{x-1}$, with $q = 1 - p$
- Note that $\partial q^x / \partial x = xq^{x-1}$ and that derivatives are linear operators

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} xpq^{x-1} = p \sum_{x=1}^{\infty} \frac{\partial q^x}{\partial x} = p \frac{\partial}{\partial x} \left(\sum_{x=1}^{\infty} q^x \right)$$

- Sum inside derivative is geometric. Sums to $q/(1-q)$

$$\mathbb{E}[X] = p \frac{\partial}{\partial x} \left(\frac{q}{1-q} \right) = \frac{p}{(1-q)^2} = \frac{1}{p}$$

- Time to first success is inverse of success probability. Reasonable

Expected value of Poisson RV

- For a Poisson RV $p(x) = e^{-\lambda}(\lambda^x/x!)$. Expected value is (First term of sum is 0, pull λ out, use $x/x! = 1/(x-1)!$)

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda$$

- Sum is Taylor's expansion of $e^{\lambda} = 1 + \lambda + \lambda^2/2! + \ldots \lambda^x/x!$

$$\mathbb{E}[X] = \lambda e^{-\lambda} e^{-\lambda} = \lambda$$

- Poisson is limit of binomial for large number of trials n with $\lambda = np$
- Counts number of successes in n trials that succeed with prob. p
- Expected number of successes is $\lambda = np$,
  - $\Rightarrow$ Number of trials $\times$ probability of individual success

Definition for continuous RVs

- Continuous RV X taking values on $\mathbb{R}$ with pdf $f(x)$
- The expected value of the RV X is

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x f(x)\, dx$$

- Compare with $\mathbb{E}[X] := \sum_{x:p(x)>0} x p(x)$ in the discrete RV case

Expected values of normal RV

- For a normal RV (add and subtract μ, separate integrals)

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x e^{\frac{(x-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x+\mu-\mu) e^{\frac{(x-\mu)^2}{2\sigma^2}} \\
&= \mu\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{\frac{(x-\mu)^2}{2\sigma^2}} + \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (x-\mu) e^{\frac{(x-\mu)^2}{2\sigma^2}}
\end{aligned}$$

- First integral is 1 because it integrates a pdf in the whole real line.
- Second integral is 0 by symmetry
- Then $\Rightarrow \mathbb{E}[X] = \mu$
- The mean of a RV with a symmetric pdf is the point of symmetry

Expected value of uniform and exponential RVs

- For a uniform RV $f(x) = 1/(b-a)$ between a and b. Expectation is

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x f(x)\,dx = \int_a^b \frac{x}{b-a}\,dx = \frac{b^2-a^2}{2(b-a)} = (a+b)/2$$

- Of course, since pdf is symmetric around $(a+b)/2$

- For an exponential RV (non symmetric) simply integrate by parts

$$\mathbb{E}[X] = \int_0^{\infty} x e^{-\lambda x}\,dx \qquad = x e^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\,dx$$

$$= x e^{-\lambda x}\Big|_0^{\infty} + \frac{e^{-\lambda x}}{\lambda}\Big|_0^{\infty} \qquad = \frac{1}{\lambda}$$

Expected value of a function of a RV

- Consider a function $g(X)$ of a RV X. Expected value of $g(X)$?
- g(X) is also a RV, then it also has a pmf $p_{g(X)}\big(g(\mathbf{X})\big)$

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

- If possible values of X are x_i possible values of $g(X)$ are $g(x_i)$ and

$$p_{g(X)}\big(g(x_i)\big) = p_X(x_i)$$

- Then we can write $\mathbb{E}\left[g(X)\right]$ as

$$\mathbb{E}\left[g(X)\right] = \sum_{i=1}^{\infty} g(x_i)p_{g(X)}\big(g(x_i)\big) = \sum_{i=1}^{\infty} g(x_i)p_X(x_i)$$

- Weighted average of functional values. No need to find pmf of $g(X)$
- Same thing can be proved for a continuous RV

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

Expected value of a linear transformation

- Consider a linear function $g(X) = aX + b$

$$\begin{aligned}
\mathbb{E}\left[aX + b\right] &= \sum_{i=1}^{\infty} (ax_i + b) p_X(x_i) \\
&= \sum_{i=1}^{\infty} a x_i p_X(x_i) + \sum_{i=1}^{\infty} b p_X(x_i) \\
&= a \sum_{i=1}^{\infty} x_i p_X(x_i) + b \sum_{i=1}^{\infty} p_X(x_i) \\
&= a\mathbb{E}\left[X\right] + b1
\end{aligned}$$

- Can interchange expectation with additive/multiplicative constants

$$\Rightarrow \mathbb{E}\left[aX + b\right] = a\mathbb{E}\left[X\right] + b$$

- Indicator function indicates an event by taking value 1 and 0 else
- Let $\mathcal{X}$ be a set $\Rightarrow \mathbb{I}\{x \in \mathcal{X}\} = 1$, if $x \in \mathcal{X}$
  $\Rightarrow \mathbb{I}\{x \in \mathcal{X}\} = 0$, if $x \notin \mathcal{X}$
- Expected value of $\mathbb{I}\{x \in \mathcal{X}\}$ (discrete case)

$$\mathbb{E}\left[\mathbb{I}\{x \in \mathcal{X}\}\right] = \sum_{x:p_X(x)>0} \mathbb{I}\{x \in \mathcal{X}\} p_X(x) = \sum_{x \in \mathcal{X}} p_X(x) = \mathrm{P}\left[x \in \mathcal{X}\right]$$

- Likewise in the continuous case

$$\mathbb{E}\left[\mathbb{I}\{x \in \mathcal{X}\}\right] = \int_{-\infty}^{\infty} \mathbb{I}\{x \in \mathcal{X}\} f_X(x) = \int_{x \in \mathcal{X}} f_X(x) = \mathrm{P}\left[x \in \mathcal{X}\right]$$

- Expected value of indicator variable = Probability of indicated event
- Compare with expectation of Bernoulli RV (it "indicates success")

Moments, central moments & variance

- n-th moment of a RV is the expected value of its n-th power $\mathbb{E}[X^n]$

$$\mathbb{E}[X^n] = \sum_{i=1}^{\infty} x_i^n p(x_i)$$

- n-th central moment corrects for expected value $\mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^n\right]$

$$\mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^n\right] = \sum_{i=1}^{\infty} \left(x_i - \mathbb{E}[X]\right)^n p(x_i)$$

- 0-th order moment is $\mathbb{E}\left[X^0\right] = 1$; 1-st moment is the mean $\mathbb{E}[X]$
- Second central moment is the variance. Measures width of the pmf

$$\text{var}[X] = \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X]$$

- 3-rd moment measures skewness (0 if pmf symmetric around mean)
- 4-th moment measures heaviness of tails (related to kurtosis)

Joint probability distributions

Sigma algebras and probability spaces

Conditional probability, independence, total probability, Bayes's rule

Random variables

Discrete random variables

- Commonly used discrete random variables

Continuous random variables

- Commonly used continuous random variables

Expected values

- Discrete random variables
- Continuous random variables
- Functions of random variables

Joint probability distributions

Joint expectations

- Independence

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

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

- 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

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

- 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

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

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

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

- 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

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

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

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

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

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

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

Joint expectations

Sigma algebras and probability spaces

Conditional probability, independence, total probability, Bayes's rule

Random variables

Discrete random variables
- Commonly used discrete random variables

Continuous random variables
- Commonly used continuous random variables

Expected values
- Discrete random variables
- Continuous random variables
- Functions of random variables

Joint probability distributions

Joint expectations
- Independence

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,

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

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

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

- Used marginal expressions

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

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)

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

- 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

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

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

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

$\square$

Variance of a sum of independent RVs

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

- Notice that mean of X is $\mathbb{E}\left[X\right] = \sum_{n=1}^{N} \mu_n$. Then

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

- Expand square and interchange summation and expectation

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

Variance of a sum of independent RVs (continued)

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

$$
\begin{aligned}
\text{var}\left[X\right] &= \sum_{n=1, {\color{red}n\neq m}}^{N} \sum_{m}^{N} \mathbb{E}\Big[(X_n - \mu_n)(X_m - \mu_m)\Big] + \sum_{n=1}^{N} \mathbb{E}\Big[(X_n - \mu_n)^2\Big] \\
&= \sum_{n=1, n\neq m}^{N} \sum_{m}^{N} \mathbb{E}(X_n - \mu_n)\mathbb{E}(X_m - \mu_m) + \sum_{n=1}^{N} \mathbb{E}\Big[(X_n - \mu_n)^2\Big] \\
&= \sum_{n=1}^{N} \text{var}\left[X_n\right]
\end{aligned}
$$

- 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

Covariance example

- Let X be a zero mean random signal and Z zero mean noise
- Signal X and noise Z are independent
- Consider received signals $Y_1 = X + Z$ and $Y_2 = -X + Z$
- Y_1 and X are positively correlated (X, Y_1 move in same direction)

$$\begin{aligned}\text{cov}(X, Y_1) &= \mathbb{E}[XY_1] - \mathbb{E}[X]\,\mathbb{E}[Y_1] \\ &= \mathbb{E}[X(X+Z)] - \mathbb{E}[X]\,\mathbb{E}[X+Z]\end{aligned}$$

- Second term is 0 ($\mathbb{E}[X] = 0$). For first term independence of X, Z

$$\mathbb{E}[X(X+Z)] = \mathbb{E}\left[X^2\right] + \mathbb{E}[X]\,\mathbb{E}[Z] = \mathbb{E}\left[X^2\right]$$

- Combining observations $\Rightarrow \text{cov}(X, Y_1) = \mathbb{E}\left[X^2\right]$

Covariance example (continued)

- Y_2 and X are negatively correlated (X, Y_1 move opposite direction)
- Same computations $\Rightarrow \text{cov}(X, Y_1) = -\mathbb{E}\left[X^2\right]$
- Can also compute correlation between Y_1 and Y_2

$$\begin{aligned}\text{cov}(Y_1, Y_2) &= \mathbb{E}\left[(X+Z)(-X+Z)\right] - \mathbb{E}\left[(X+Z)\right]\mathbb{E}\left[(-X+Z)\right] \\ &= -\mathbb{E}\left[X^2\right] + \mathbb{E}\left[Z^2\right]\end{aligned}$$

- Negative correlation if $\mathbb{E}\left[X^2\right] > \mathbb{E}\left[Z^2\right]$ (small noise)
- Positive correlation if $\mathbb{E}\left[X^2\right] < \mathbb{E}\left[Z^2\right]$ (large noise)
- Correlation between X and Y_1 or X and Y_2 comes from causality
- Correlation between Y_1 and Y_2 does not
- Plausible, indeed commonly used, model of a communication channel