# Chapter 3 Part 2: Random variables

Jess Kunke

MATH/STAT 394: Probability I (Summer 2022 A-term)

# Outline

Mid-course feedback, midterm example

Wrap up cdfs (with practice)

(Great) Expectations

Variance

Median and quantiles

# Outline

Mid-course feedback, midterm example

Wrap up cdfs (with practice)

(Great) Expectations

Variance

Median and quantiles

# Mid-course feedback

Thank you for the feedback!

- Most said pace is fast
  - Definitely! Accelerated course is very fast
  - I will try to speak more slowly, leave slides up longer
- Most said homework difficulty is fine
- Study groups have been helpful
- Connecting new problems to ones we've already seen has been helpful

# Midterm: updated time

- This Friday July 8th, 9-10am, CMU 230
- Bring one or more pens/pencils and a half-sheet of paper with whatever handwritten notes you'd like
- I will provide blank paper and the exam instructions
- Not all problems are equal length

# Midterm work example

Two fair dice are rolled. What is the conditional probability that at least one lands on 6 given that the dice land on different numbers?

Fine:

$$
\begin{aligned}
P(\text{at least one } 6 \mid \text{different}) &= \frac{P(\text{at least one } 6, \text{different})}{P(\text{different})} \\
&= \frac{P\{1st = 6, 2nd \neq 6\} + P\{1st \neq 6, 2nd = 6\}}{5/6} \\
&= \frac{(1/6) \cdot (5/6) + (1/6) \cdot (5/6)}{5/6} \\
&= \frac{1}{3}. \quad \text{(this step not necessary unless specified)}
\end{aligned}
$$

# Midterm work example

Two fair dice are rolled. What is the conditional probability that at least one lands on 6 given that the dice land on different numbers?

Not enough:

A = {at least one 6}

B =

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$
$$= \frac{1}{3}.$$

not enough steps

# Outline

Mid-course feedback, midterm example

Wrap up cdfs (with practice)

(Great) Expectations

Variance

Median and quantiles

# Key integrals and derivatives

$\beta_0 + \beta_1 x + \beta_2 x^2 + \dots$

- Polynomials $\frac{d}{dx} x = 1 \quad \frac{d}{dx} x^2 = 2x$

$$\frac{d}{dx} a x^k = a k x^{k-1}$$

$$\int x^3 dx = ? \qquad \frac{d}{dx} x^4 = 4x^3 \Rightarrow \int_a^b x^3 dx = \left[\frac{1}{4} x^4\right]_{x=a}^{x=b}$$

- Exponential

$$\frac{d}{dx} e^{ax} = a e^{ax} \qquad \int_m^b e^{ax} dx = \left[\frac{1}{a} e^{ax}\right]_{x=m}^{x=b}$$

## pmfs, pdfs, and cdfs

Discrete RVs

- Probability mass function (pmf)

$$p(k) = P(X = k) \quad \text{for all possible values } k \text{ of } X$$

- Cumulative distribution function (cdf)

$$F(s) = P(X \leq s) = \sum_{k:\, k \leq s} P(X = k)$$

Continuous RVs

- Cumulative distribution function (cdf)

$$F(s) = P(X \leq s) = \int_{-\infty}^{s} f(x)dx \quad \text{for all } s \in \mathbb{R}$$

- Probability density function (pdf)

not a probability

$$f \text{ such that } P(X \leq s) = \int_{-\infty}^{s} f(x)dx \quad \text{for all } s \in \mathbb{R}$$

Other RVs

- Cumulative distribution function (cdf)

$$F(s) = P(X \leq s) \quad \text{for all } s \in \mathbb{R}$$

# Do discrete and continuous RVs partition the space of possible RVs?

- If $F$ is piece-wise constant
  $\implies$ it is the cdf of a **discrete RV**
- If $F$ is continuous
  $\implies$ it is the cdf of a **continuous RV**
- If $F$ is discontinuous and not piece-wise constant
  $\implies$ neither discrete nor continuous RV
  - but we can still compute probabilities using the cdf
  - e.g. mixtures of distributions

The cdf exists for **any** RV

# Try at home

- Go through the examples we covered in lecture last time
- Pick some of the simpler distributions we've covered (flipping a coin, rolling a fair die, binomial, uniform, exponential)
  - Graph and write the pmf/pdf and cdf
  - Do it in whatever order makes sense to you, then try doing it in a different order
  - How could you tell the cdf from the pmf/pdf?
  - How could you tell the pmf/pdf from the cdf?

# Why did we introduce the cdf?

**Theoretical reason**

- We only need $P(X \leq t)$ for any $t$ to compute any prob. measure
- Therefore the cdf is sufficient for our purposes

# Why did we introduce the cdf?

**Theoretical reason**

- We only need $P(X \leq t)$ for any $t$ to compute any prob. measure
- Therefore the cdf is sufficient for our purposes

**Practical reason**

- The cdf itself is a prob. so we can use classical rules of prob. to manipulate it
- On the other hand the pdf is just a function and sometimes it is not practical or does not exist

# Why did we introduce the cdf?

complement
$1 - $
$P(M>t) \rightarrow$ cdf $\xrightarrow{\text{deriv.}}$ pdf
$P(M \le t)$

**Example**

Let $X \sim \text{Expo}(\lambda)$, $Y \sim \text{Expo}(\mu)$ be independent.
What is the pdf of $M = \min(X, Y)$?
Recall that $F_X(t) = 1 - e^{-\lambda t}$. $= P(X \le t)$ $\quad P(X>t) = 1 - P(X \le t) = e^{-\lambda t}$.
cdf of an exponentially distributed RV.

Tips:

- Notice that event $\{\min\{X, Y\} > t\}$ is equivalent to $\{X > t\} \cap \{Y > t\}$
- Find the cdf of $M$, then use it to find the pdf
- What is the name of the distribution of $M$? $\rightarrow M \sim$ ___ $(\quad)$

$$P(M>t) = P(\min(X,Y)>t) = P(X>t, Y>t)$$
$$= P(X>t)P(Y>t) \qquad X, Y \text{ independent}$$
$$= e^{-\lambda t} e^{-\mu t}$$
$$= e^{-(\lambda+\mu)t} \implies M \sim \text{Expo}(\lambda+\mu)$$

$F_M(t) = P(M \le t) = 1 - e^{-(\lambda+\mu)t}$

$\Rightarrow f_M(t) = \frac{d}{dt} P(M \le t) = (\lambda+\mu)e^{-(\lambda+\mu)t}$ (could also write $F_M'(t)$)

# Tips on probability distributions

Wikipedia pages on probability distributions are a great resource!

- Check out the distributions from class (binomial, uniform, exponential, etc.)
- Shows pmf/pdf, cdf, and lots of other properties
- Presents definitions and applications, connections to other distributions, and sometimes some history
- You can explore some new distributions you haven't seen before too

# Outline

Mid-course feedback, midterm example

Wrap up cdfs (with practice)

(Great) Expectations

Variance

Median and quantiles

# Expectation

Motivation

- Given a RV, we have numerous tools to compute probabilities
- We said that we also sometimes want to know what kind of result we expect on average, a "typical value" for a given RV
- e.g. if you flip a coin n times, what is the average number of tails you should get?
- In probability, this "average" number is called an **expectation** and it is a central object

# Intuition

## Example

At a casino, suppose

- you lose 1$ 90% of the time,
- you gain 10$ 9% of the time, and
- you gain 100$ 1% of the time.

What is your expected net gain?

# Intuition

### Example

At a casino, suppose

- you lose 1$ 90% of the time,
- you gain 10$ 9% of the time, and
- you gain 100$ 1% of the time.

What is your expected net gain?

- First understand that the average is a **number** not a probability
- Then

$$\text{expected net gain} = \underbrace{(-1)}_{\text{net gain}} \cdot \underbrace{\frac{90}{100}}_{\text{frequency}} + 10 \cdot \frac{9}{100} + 100 \cdot \frac{1}{100} = 1$$

# Expectation of a discrete RV

## Definition

The **expectation** or **mean** of a discrete random variable $Y$ is defined by

$$E(Y) = \sum_k kP(X = k).$$

Expectation is often written with square brackets, $E[Y]$.

## Example

What is the expectation of $X \sim \mathrm{Ber}(p)$?

$$P(X=k) = \begin{cases} p & k=1 \\ 1-p & k=0 \end{cases}$$

$$E(X) = 0\cdot(1-p) + 1\cdot p = p.$$

# Expectation of a discrete RV

## Definition

The **expectation** or **mean** of a discrete random variable $Y$ is defined by

$$E(Y) = \sum_k kP(X = k).$$

Expectation is often written with square brackets, $E[Y]$.

## Example

What is the expectation of $X \sim \text{Ber}(p)$? → $E[X] = p$.

**Link between expectation and probability**

- For an event $A \subseteq \Omega$ the **indicator RV** of $A$ (denoted $\mathbb{1}_A(\omega)$ or $I_A(\omega)$) is

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

- $\mathbb{1}_A \sim \text{Ber}(P(A))$ (since $P(\mathbb{1}_A = 1) = P(\omega \in A)$) $E[\mathbb{1}\{\omega \in A\}] = P(A)$
- Therefore

$$\text{E}[\mathbb{1}_A] = P(A).$$

# Expectation of a continuous RV

For continuous random variables, we replace the sum over the pmf with an integral over the pdf:

### Definition

Suppose $Y$ is a continuous random variable with pdf $f$. Then the **expectation** or **mean** of $Y$ (often denoted $\mu_Y$) is defined by

$$E[Y] = \int_{-\infty}^{\infty} y f(y) dy.$$

### Example Unif[a, b]

Let $X \sim$ ~~[a, b]~~. Find E[$X$].

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \left[ \frac{x^2}{2(b-a)} \right]_{x=a}^{x=b}$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}.$$

## Comments on expectations

The cases we've seen had finite expectations.

- Expectation can be infinite or undefined (see book examples)

# Comments on expectations

- Expectation can be infinite or undefined (see book examples)
- Expectation can be seen as the "center of mass" of the distribution

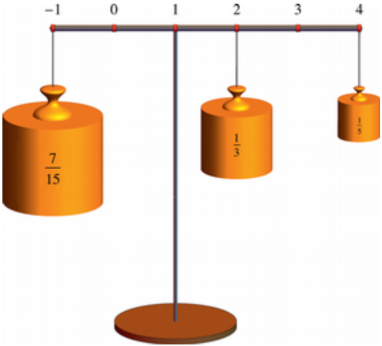



Figure: Figure 3.8 from the textbook

# Expectation of a function of a RV

If we know the distribution of a RV $X$ and now we are interested in a RV $Y = g(X)$ for some function $g$, do we have to compute the distribution and expectation from scratch? No.

## Theorem

*Let $X$ be a RV that takes values in $\mathcal{X}$ and $g : \mathcal{X} \to \mathbb{R}$ be some function.*

$$E[g(X)] = \sum_{k \in \mathcal{X}} g(k)p(k) \qquad \text{if } X \text{ is discrete with pmf } p,$$

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx \qquad \text{if } X \text{ is continuous with pdf } f.$$

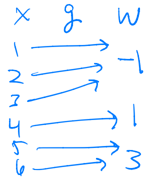
say $X$ = die roll $= \begin{cases} 1 \\ \vdots \\ 6 \end{cases}$ with prob. 1/6

now we want $W = \begin{cases} -1 & X = 1,2,3 \\ 1 & X = 4 \\ 3 & X = 5,6 \end{cases}$

Proof of discrete case:

Notice that $P(W=-1) = P(X=1) + P(X=2) + P(X=3)$

$$E[g(x)] = \sum_y y P(g(x) = y)$$




Later, we will cover how to derive the distribution of $Y$ from the dist. of $X$

$$E[g(X)] = \sum_y y\, P(g(X) = y)$$

$$= \sum_y \sum_{k:\, g(k)=y} y\, P(X=k)$$

$$= \sum_y \sum_{k:\, g(k)=y} g(k)\, P(X=k)$$

$$= \sum_{k \in X} g(k)\, P(X=k).$$

# Linearity of expectation

## Theorem

1. *For any random variable* $X$ *and any* $a, b \in \mathbb{R}$, $E[aX + b] = aE[X] + b$.
2. *If* $X, Y$ *are random variables on the same probability space, then* $E[X + Y] = E[X] + E[Y]$.
3. *Let* $X_1, ..., X_n$ *be* $n$ *random variables defined on the same probability space and* $g_1, ..., g_n$ *be* $n$ *functions. Then*

$$E[g_1(X_1) + ... + g_n(X_n)] = E[g_1(X_1)] + ... + E[g_n(X_n)].$$

# Linearity of expectation

## Theorem

1. *For any random variable $X$ and any $a, b \in \mathbb{R}$, $E[aX + b] = aE[X] + b$.*
2. *If $X, Y$ are random variables on the same probability space, then $E[X + Y] = E[X] + E[Y]$.*
3. *Let $X_1, ..., X_n$ be $n$ random variables defined on the same probability space and $g_1, ..., g_n$ be $n$ functions. Then*

$$E[g_1(X_1) + ... + g_n(X_n)] = E[g_1(X_1)] + ... + E[g_n(X_n)].$$

## Example

Using the linearity of expectation, compute the expectation of $X \sim \text{Bin}(n, p)$.

$X \sim Bin(n,p) \Rightarrow X = \sum_{i=1}^{n} Y_i$, $Y_i \overset{iid}{\sim} Ber(p)$. We know $E[Y_i] = p$.

$E[X] = E\left[\sum_{i=1}^{n} Y_i\right]$

$= \sum_{i=1}^{n} E[Y_i]$ linearity of exp.

$= np.$

**Example**

Anne has three 4-sided dice, two 6-sided dice and one 12-sided die. All the dice are fair and numbered 1, 2, ..., $n$ for $n = 4, 6$, or 12. She rolls all the dice and adds up the numbers showing. What is the expected value of the sum?

Let $A_1, A_2, A_3$ represent the numbers showing on the 4-sided dice

$B_1, B_2$ represent the rolls of the 6-sided dice

$C_1$ for the 12-sided die.

Let $X = A_1 + A_2 + A_3 + B_1 + B_2 + C_1$.

$E[A_i] = 2.5$ $\quad$ $E[B_i] = 3.5$ $\quad$ $E[C_i] = 6.5$

$(1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4})$

1 2 3 4

$\frac{1+2+3+4}{4}$

By lin. of exp.,

$$E[X] = 3E[A_i] + 2E[B_i] + E[C_i]$$
$$= 3(2.5) + 2(3.5) + 6.5$$
$$= 21.$$

# Outline

# Variance

**Motivation**

- The expectation summarizes the RV to a single point
- Generally the distribution should gather around the mean, but how much?
- The variance informs us about the **dispersion** of the RV around the mean

# Variance

## Definition

The **variance** of a random variable $X$ with mean $\mu$ is defined as

$$\mathrm{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right]$$

$\mathrm{Var}(X)$ is often denoted $\sigma_X^2$.

# Variance

## Definition

The **variance** of a random variable $X$ with mean $\mu$ is defined as

$$\mathrm{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right]$$

$\mathrm{Var}(X)$ is often denoted $\sigma_X^2$.
The square root $\sigma_X$ of the variance is called the **standard deviation**.

# Variance

## Definition

The **variance** of a random variable $X$ with mean $\mu$ is defined as

$$\mathrm{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right]$$

$\mathrm{Var}(X)$ is often denoted $\sigma_X^2$.
The square root $\sigma_X$ of the variance is called the **standard deviation**.
In terms of pmf or pdf, we have that

$$\mathrm{Var}(X) = \sum_{k \in \mathcal{X}} (x-\mu)^2 p(k) \qquad \text{for a discrete RV with pmf } p,$$

$$\mathrm{Var}(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx \qquad \text{for a continuous RV with pdf } f.$$

# Variance

## Definition

The **variance** of a random variable $X$ with mean $\mu$ is defined as

E[X]

$$\mathrm{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right] \quad = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right]$$

$\mathrm{Var}(X)$ is often denoted $\sigma_X^2$.
The square root $\sigma_X$ of the variance is called the **standard deviation**.
In terms of pmf or pdf, we have that

$$\mathrm{Var}(X) = \sum_{k \in \mathcal{X}} (x-\mu)^2 p(k) \qquad \text{for a discrete RV with pmf } p,$$

$$\mathrm{Var}(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx \qquad \text{for a continuous RV with pdf } f.$$

Note:

- Variance is defined through the expectation of a function of the RV
- This is true of many characteristics of a RV: expectation is our main tool
- As with expectation, the variance may be finite, infinite or undefined

# Variance

## Definition

The **variance** of a random variable $X$ with mean $\mu$ is defined as

$$\mathrm{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right]$$

$\mathrm{Var}(X)$ is often denoted $\sigma_X^2$.
The square root $\sigma_X$ of the variance is called the **standard deviation**.
In terms of pmf or pdf, we have that

$$\mathrm{Var}(X) = \sum_{k \in \mathcal{X}} (x-\mu)^2 p(k) \qquad \text{for a discrete RV with pmf } p,$$

$$\mathrm{Var}(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx \qquad \text{for a continuous RV with pdf } f.$$

## Example

If $X \sim \mathrm{Ber}(p)$, what is $\mathrm{Var}(X)$? $E[X] = p$

$$\mathrm{Var}(X) = E\left[(X-p)^2\right] = (0-p)^2 \cdot (1-p) + (1-p)^2 p$$

we used the exp. of $g(x) = (x-p)^2$.

$$= p^2 - p^3 + p - 2p^2 + p^3$$

$$= p - p^2 = p(1-p).$$

## Variance: another way to compute

Sometimes this is an easier way to compute the variance:

$E[3] = 3.$

**Lemma**
*The variance of a RV $X$ can also be expressed as*

$$Var(X) = E[X^2] - E[X]^2.$$

Proof:

$$
\begin{aligned}
Var(X) &= E[(X - E[X])^2] \\
&= E[X^2 - 2X E[X] + E[X]^2] \\
&= E[X^2] - 2E[X]E[X] + E[X]^2 \quad \text{lin. of exp.} \\
&= E[X^2] - 2(E[X])^2 + (E[X])^2 \\
&= E[X^2] - (E[X])^2.
\end{aligned}
$$

# Variance: example

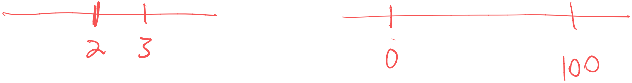
**Example**

we already have $E[X] = \frac{a+b}{2}$.

Let $X \sim \text{Unif}(a, b)$ with $a < b$. What do you think should happen to the variance as the width of the interval increases? Find $\text{Var}(X)$; does that happen in your solution?

$b-a$

$X \sim \text{Unif}(2,3)$ $\quad \text{Var}(X)$ $\qquad Y \sim \text{Unif}(0,100)$ $\quad \text{Var}(Y)$




$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_a^b \frac{x^2}{b-a}\,dx = \left[\frac{x^3}{3(b-a)}\right]_{x=a}^{x=b} = \frac{b^3-a^3}{3(b-a)}$$

$$= \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{b^2+ab+a^2}{3}.$$

$$\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4}$$

$$\frac{(a+b)^2}{4} = \frac{a^2+2ab+b^2}{4}$$

$$= \frac{b^2}{12} + \frac{a^2}{12} - \frac{2ab}{12} = \frac{(b-a)^2}{12}.$$

standard deviation $\sigma_X = \frac{b-a}{\sqrt{12}}$

# Moments

### Definition

The **nth moment** of a RV $X$ is

$$E[X^n].$$

The **nth centered moment** of a RV $X$ is

$$E[(X - E[X])^n]$$

$E[X], \; E[X^2]$

$E[(X - E[X])^2]$

# Moments

## Definition

The **nth moment** of a RV $X$ is

$$\mathrm{E}[X^n].$$

The **nth centered moment** of a RV $X$ is

$$\mathrm{E}[(X - \mathrm{E}[X])^n]$$

Notes

- 1st moment: mean
- 2nd moment: mean square
- 2nd centered moment: variance
- 3rd centered moment: kurtosis
  - Tells us about asymmetry of RV
  - 0 if RV is symmetric
- Moments are explored in more detail in MATH/STAT 395

# Variance properties

$2X \in (0, 8)$

Motivation

- Variance is **not linear**! Instead we have the following property:

$X \sim Unif(0,4)$ $\quad$ $X \sim Unif(10,14)$


## Lemma

*For a RV $X$ and $a, b \in \mathbb{R}$,*

$$Var(aX + b) = a^2 Var(X).$$

Proof: $Var(aX+b) = E\left[(aX+b - E[aX+b])^2\right]$

$$= E\left[(aX + b - aE[X] - b)^2\right]$$

$$= E\left[a^2 (X - E[X])^2\right]$$

$$= a^2 E\left[(X - E[X])^2\right] = a^2 Var(X). \quad //.$$

Takeaways:

- Adding a constant to the RV does not change the variance
- $\sigma_{aX+b} = \sqrt{Var(aX + b)} = a\sigma_X$
- Standard deviation $\sigma$ has the same 'units' as the RV or the mean, while variance $\sigma^2$ has squared units

# Null variance

Motivation

- The following theorem formalizes the intuition that if a RV does not vary (i.e. $\text{Var}(X) = 0$) then it must be a constant

### Theorem

*For a RV $X$, $Var(X) = 0$ if and only if $P(X = a) = 1$ for some constant $a \in \mathbb{R}$.*

Proof: $\Longleftrightarrow$

1) $\Longleftarrow$ :

$P(X=a) = 1 \Rightarrow E[X] = a, \quad Var(X) = 0.$

2) $\Longrightarrow$ : (discrete)

$Var(X) = 0 \Rightarrow \sum_k \underbrace{(k-\mu)^2 P(X=k)}_{\geq 0} = 0$ (where $\mu = E[X]$)

$\Rightarrow (k-\mu)^2 P(X=k) = 0 \quad \forall k.$

$\Rightarrow k = \mu$ or $P(X=k) = 0$. for each $k$.

$\Rightarrow P(X=k) > 0$ for only one value of $k$, $k = \mu$. //

$\Uparrow$

$P(X=\mu) = 1.$

# Expectation of product of independent RVs

Remember:

- $X_1, \ldots, X_n$ are independent if for any (Borel) sets $B_1, \ldots, B_n \in \mathbb{R}$,

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = P(X_1 \in B_1) \ldots P(X_n \in B_n).$$

- For an indicator RV, $\mathrm{E}[\mathbb{1}_A] = P(A)$ for $A \subseteq \Omega$

# Expectation of product of independent RVs

Remember:

- $X_1, \dots, X_n$ are independent if for any (Borel) sets $B_1, \dots, B_n \in \mathbb{R}$,

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n).$$

- For an indicator RV, $\mathrm{E}[\mathbb{1}_A] = P(A)$ for $A \subseteq \Omega$

**Another characterization of independent RV:**

- Denote $h_i(x_i) = \begin{cases} 1 & \text{if } x \in B_i \\ 0 & \text{if } x \notin B_i \end{cases}$
- Note that $h_1(x_1) \dots h_n(x_n) = \begin{cases} 1 & \text{if } x_1 \in B_1, \dots, x_n \in B_n \\ 0 & \text{otherwise} \end{cases}$
- Previous definition can be written as

$$\mathrm{E}[h_1(X_1) \dots h_n(X_n)] = \mathrm{E}[h_1(X_1)] \dots \mathrm{E}[h_n(X_n)].$$

- Namely, for independent $X_1, \dots, X_n$, the expectation of a product of some functions of RV is equal to the product of the expectation.

# Expectation of product of independent RV

Motivation:
As any function can be decomposed/approximated by indicator RVs, we get the following theorem:

**Theorem**
*$X_1, \ldots, X_n$ are independent if and only if for any functions $h_1, \ldots, h_n$,*

$$E[h_1(X_1)\ldots, h_n(X_n)] = E[h_1(X_1)]\ldots E[h_n(X_n)].$$

**Corollary**
*If $X, Y$ are independent, then*

$$Var(X+Y) = Var(X) + Var(Y).$$

Questions:

- If $X, Y, Z$ are independent, is $\mathrm{E}[XYZ] = \mathrm{E}[X]\mathrm{E}[Y]\mathrm{E}[Z]$? Yes
- If $\mathrm{E}[XYZ] = \mathrm{E}[X]\mathrm{E}[Y]\mathrm{E}[Z]$, are $X, Y, Z$ independent? Not necessarily

## Variance of independent RV

The variance result can be generalized as follows.

**Theorem**

*If $X_1, \ldots, X_n$ are independent, then*

$$Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n).$$

**Example**

What is the variance of $X \sim \text{Bin}(n, p)$? $\Rightarrow X = \sum_{i=1}^{n} Y_i = Y_1 + Y_2 + \ldots + Y_n$

$Y_i \overset{iid}{\sim} Ber(p)$.

$Var(Y_i) = p(1-p)$.

$Var X = \sum_{i=1}^{n} Var(Y_i)$ ($Y_i$ indep.)

$= \sum_{i=1}^{n} p(1-p)$

$= np(1-p)$.

# Outline

Mid-course feedback, midterm example

Wrap up cdfs (with practice)

(Great) Expectations

Variance

Median and quantiles

# Median

**Motivation**

- The expectation often gives a good summary of a RV
- Yet, if the RV has some abnormally large values, the expectation may be a bad indicator of where the center of the distribution lies
- Another indicator is often used: the median that tells us where to split the distribution of $X$ to have equal mass on the left and right sides of the median

# Median of a continuous RV

### Definition

The **median** of a continuous RV $X$ is a value $m$ s.t.

$$P(X \geq m) = P(X \leq m) = 1/2$$

# Median of a continuous RV

### Definition

The **median** of a continuous RV $X$ is a value $m$ s.t.

$$P(X \geq m) = P(X \leq m) = 1/2$$

### Example

At a call center, a phone call arrives on average every 5 min (model it as an exponential RV). What is the median time to wait for a call?

- The pdf is $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ and 0 otherwise with $\lambda = 1/5$ (since $\mathrm{E}[X] = 1/\lambda = 5$).
- To compute the median, it suffices to use the cdf. We want $m$ such that $F_X(m) = 1/2$.
- Since $F_X(t) = e^{-\lambda t}$, we get that $m = -\log(1/2)/\lambda \approx 3.47$.

# Median of discrete RV

$p(X=-1) = p(X=0) = 1/3$

## Example

Consider $X$ uniformly distributed on $\{-1, 0, 1\}$ (discrete uniform).
How can we define a median for $X$?

- Here there does not exist $m$ s.t. $P(X \leq m) = P(X \geq m) = 1/2$.
- For example $P(X \leq 0) = 2/3$ and $P(X \geq 0) = 2/3$.
- The problem is that here 0 takes some probability mass so we need to slightly change the definition of a median in the discrete case

# Median of discrete RV

### Example

Consider $X$ uniformly distributed on $\{-1, 0, 1\}$ (discrete uniform).
How can we define a median for $X$?

- Here there does not exist $m$ s.t. $P(X \leq m) = P(X \geq m) = 1/2$.
- For example $P(X \leq 0) = 2/3$ and $P(X \geq 0) = 2/3$.
- The problem is that here 0 takes some probability mass so we need to slightly change the definition of a median in the discrete case

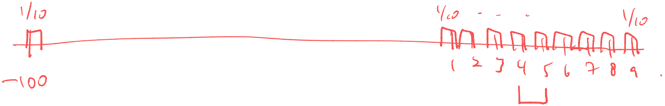
### Definition

Generally, a **median** of a RV $X$ is any value $m$ such that

$$P(X \geq m) \geq 1/2 \qquad P(X \leq m) \geq 1/2$$

So in the above example, 0 would be a median.

# Median




## Example

Let $X$ be uniformly distributed on $\{-100, 1, 2, 3, \ldots 9\}$. So $X$ has a prob. dist.

$$P(X = -100) = 1/10, \qquad P(X = k) = 1/10 \quad \text{for } k \in \{1, \ldots 9\}$$

What are the expectation and the median of $X$?

- $\mathrm{E}[X] = -100 \cdot 1/10 + (1 + 2 + \ldots + 9) \cdot 1/10 = -5.5$
- On the other hand,

$$P(X \leq 4.5) = p(-100) + p(1) + p(2) + p(3) + p(4) = 1/2$$
$$P(X \geq 4.5) = p(5) + \ldots + p(9) = 1/2$$

- So 4.5 is **a** median for $X$
- Any $m \in [4, 5]$ is a median for $X$; we usually take the mid-point of the interval
- A median (e.g. 4.5) illustrates much better than the mean (-5.5) the fact that 90% of the possible values are in $\{1, \ldots, 9\}$
- The mean better represents the center of (probability) mass

# Quantiles

**Motivation**

- Let's generalize the median
- Typically we would like to know if some observation of our RV is **rare** or not
- Namely we would like to have access to a value $x$, such that if $X \geq x$ then the probability of this observation is small
- This is formalized with the definitions of **quantiles**

# Quantiles

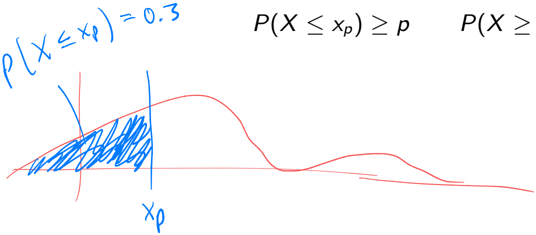
### Definition

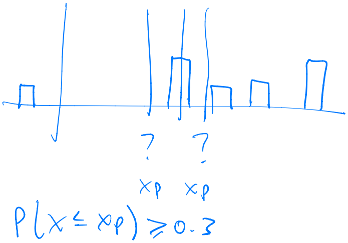
Given $0 \le p \le 1$ (e.g. $p = 90/100$), the $\mathbf{p^{th}}$ **quantile** of a continuous RV $X$ is any value $x_p$ such that

$$P(X \le x_p) = p \qquad P(X \ge x_p) = 1 - p$$

More generally the $\mathbf{p^{th}}$ **quantile** of a RV $X$ is any value $x_p$ such that

$$P(X \le x_p) \ge p \qquad P(X \ge x_p) \ge 1 - p.$$



# Quantiles

## Definition

Given $0 \leq p \leq 1$ (e.g. $p = 90/100$), the **p$^{\text{th}}$ quantile** of a continuous RV $X$ is any value $x_p$ such that

$$P(X \leq x_p) = p \qquad P(X \geq x_p) = 1 - p$$

More generally the **p$^{\text{th}}$ quantile** of a RV $X$ is any value $x_p$ such that

$$P(X \leq x_p) \geq p \qquad P(X \geq x_p) \geq 1 - p.$$

**Notes**

- $p = 1/2$: we retrieve the median! (i.e. median = 0.5th quantile or 50th percentile) 0.9th quant. 90th perc.
- $p = 90/100$: the 90$^{\text{th}}$ quantile tells us that there is less than 10% chance of observing a value greater than $x_p$
- In the second definition, we want to take into account values of $x_p$ that could have a non-zero mass but still satisfy the idea of a quantile.

# Quantiles

## Definition

Given $0 \leq p \leq 1$ (e.g. $p = 90/100$), the $\mathbf{p^{th}}$ **quantile** of a continuous RV $X$ is any value $x_p$ such that

$$P(X \leq x_p) = p \qquad P(X \geq x_p) = 1 - p$$

More generally the $\mathbf{p^{th}}$ **quantile** of a RV $X$ is any value $x_p$ such that

$$P(X \leq x_p) \geq p \qquad P(X \geq x_p) \geq 1 - p.$$

**Notes**

- $p = 1/2$: we retrieve the median! (i.e. median = 0.5th quantile or 50th percentile)
- $p = 90/100$: the $90^{\text{th}}$ quantile tells us that there is less than 10% chance of observing a value greater than $x_p$
- In the second definition, we want to take into account values of $x_p$ that could have a non-zero mass but still satisfy the idea of a quantile.

# Quantiles

## Definition

Given $0 \leq p \leq 1$ (e.g. $p = 90/100$), the $\mathbf{p^{th}}$ **quantile** of a continuous RV $X$ is any value $x_p$ such that

$$P(X \leq x_p) = p \qquad P(X \geq x_p) = 1 - p$$

More generally the $\mathbf{p^{th}}$ **quantile** of a RV $X$ is any value $x_p$ such that

$$P(X \leq x_p) \geq p \qquad P(X \geq x_p) \geq 1 - p.$$

**Notes**

- $p = 1/2$: we retrieve the median! (i.e. median = 0.5th quantile or 50th percentile)
- $p = 90/100$: the $90^{th}$ quantile tells us that there is less than 10% chance of observing a value greater than $x_p$
- In the second definition, we want to take into account values of $x_p$ that could have a non-zero mass but still satisfy the idea of a quantile.

# Quantiles

## Definition

Given $0 \leq p \leq 1$ (e.g. $p = 90/100$), the $\mathbf{p^{th}}$ **quantile** of a continuous RV $X$ is any value $x_p$ such that

$$P(X \leq x_p) = p \qquad P(X \geq x_p) = 1 - p$$

More generally the $\mathbf{p^{th}}$ **quantile** of a RV $X$ is any value $x_p$ such that

$$P(X \leq x_p) \geq p \qquad P(X \geq x_p) \geq 1 - p.$$

**Notes**

- $p = 1/2$: we retrieve the median! (i.e. median $=$ 0.5th quantile or 50th percentile)
- $p = 90/100$: the $90^{th}$ quantile tells us that there is less than 10% chance of observing a value greater than $x_p$
- In the second definition, we want to take into account values of $x_p$ that could have a non-zero mass but still satisfy the idea of a quantile.