

Math 4261: Mathematical Statistics I

Lecture 2: Probability review I

Single variable tools

Ref: Blitzstein-Hwang

Fall 2026

Outline

Random variables

PMF

PDF

CDFs and quantiles

Expectation

Variance

Deviation inequalities

Outline

Random variables

PMF

PDF

CDFs and quantiles

Expectation

Variance

Deviation inequalities

Recap: Polling example

In our Harris v. Trump polling, we found that for either sampling scheme

$$X_i \sim \text{Bern}(\theta) \quad \text{for every } i \text{ (marginally),} \quad \hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i.$$

The marginal distributions are the same, but the dependence is different:

$$n\hat{\theta} = \sum_{i=1}^n X_i \sim \begin{cases} \text{Binomial}(n, \theta), & \text{with replacement,} \\ \text{Hypergeometric}(N, N\theta, n), & \text{without replacement.} \end{cases}$$

Matching marginals do not imply matching joint distributions.

Today we review some marginal concepts: the behavior of a *single* random variable. Next lecture, we return to joint distributions and dependence.

Random variables

- **Discrete r.v.'s.** X takes finite or countably infinite many values.

E.g. X = number of heads in 100 coin flips,

X = number of die rolls to first 6.

Special case: **Indicator r.v.'s.** X takes only two values: 0 or 1. May encode yes/no outcomes. E.g. Toss a die. Set

$$X = \begin{cases} 1 & \text{if roll an even number} \\ 0 & \text{if roll an odd number} \end{cases}$$

- **Continuous r.v.'s.** X takes a continuum of values.

E.g. X = position of a randomly spun wheel.

The *distribution* or *law* of X describes $\mathbb{P}[X \in A]$ for any set $A \subseteq \mathbb{R}$.

Probability mass functions

Let X be a discrete r.v. taking values in a set S . The **probability mass function** (PMF) of X is a function $p_X : S \rightarrow [0, 1]$ defined by

$$p_X(x) = \mathbb{P}[X = x], \text{ for all } x \in S$$

The set S is called the **support** of X .

Properties:

- (i) $p_X(x) > 0$ for $x \in S$, $p_X(x) = 0$ otherwise.
- (ii) $\sum_{x \in S} p_X(x) = 1$

Bernoulli random variables

An r.v. X has the **Bernoulli** distribution with **success probability** $p \in [0, 1]$ if

$$\mathbb{P}[X = 1] = p \quad \text{and} \quad \mathbb{P}[X = 0] = 1 - p.$$

- ▶ We write $X \sim \text{Bern}(p)$.
- ▶ Any $\{0, 1\}$ -valued r.v. has a Bernoulli distribution.
- ▶ The indicator r.v. $\mathbf{1}_A$ of some event A is a Bernoulli r.v. with $\mathbf{1}_A \sim \text{Bern}(p)$ where $p = \mathbb{P}[A]$.
- ▶ Bernoulli r.v.'s model the outcomes of **Bernoulli trials** which are experiments with a “success” or “failure” outcome.

Binomial random variables

An r.v. X has the **Binomial** distribution with n trials and success probability $p \in [0, 1]$ if

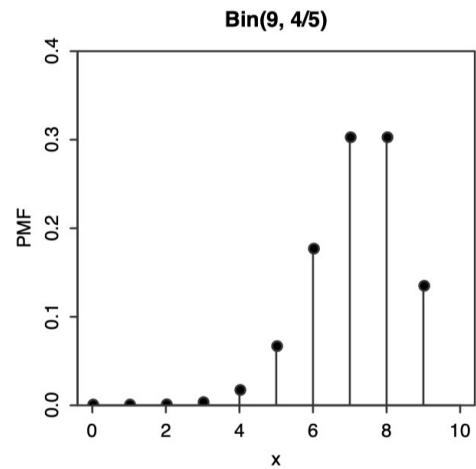
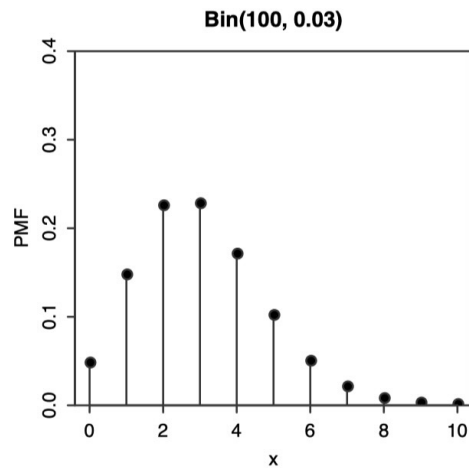
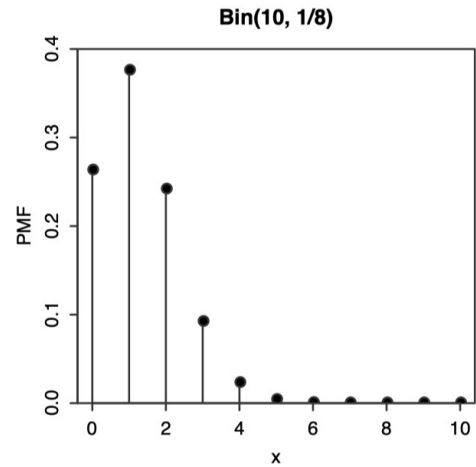
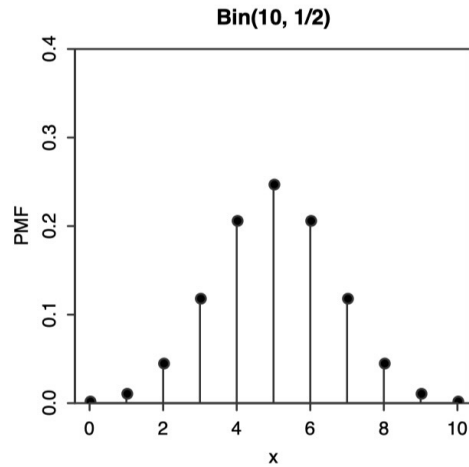
X = number of successes in n independent $\text{Bern}(p)$ trials

- ▶ We write $X \sim \text{Bin}(n, p)$.
- ▶ Equivalently, $X = \sum_{i=1}^n X_i$, where $X_i \sim \text{Bern}(p)$ are independent.
- ▶ Notice $\text{Bin}(1, p) = \text{Bern}(p)$.

PMF:

$$\mathbb{P}[X = k] = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}, \quad \text{for } k = 0, 1, \dots, n$$

Binomial plots



Poisson

X has the **Poisson distribution with rate λ** if X takes values in $\{0, 1, 2, 3, \dots\}$ with PMF

$$\mathbb{P}[X = k] = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{for } k \geq 0.$$

Write $X \sim \text{Pois}(\lambda)$.



Siméon Denis Poisson

Poisson approximation of Binomial

Suppose $X_n \sim \text{Bin}\left(n, \frac{\lambda}{n}\right)$. Then

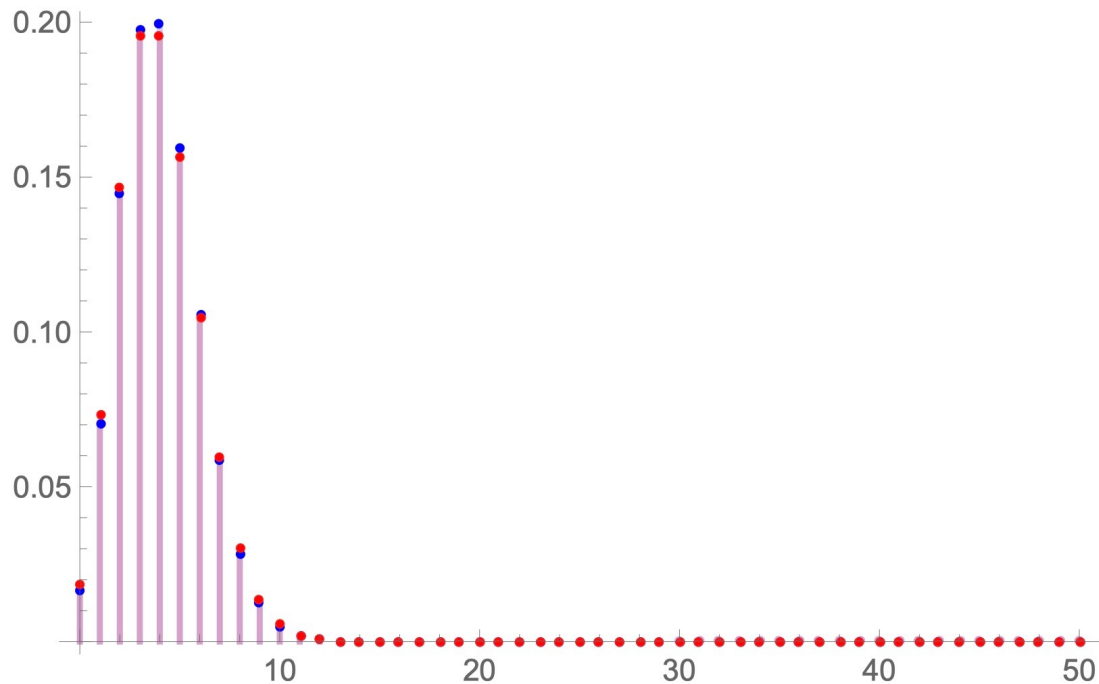
$$X_n \sim \text{Pois}(\lambda) \quad \text{approximately}$$

More precisely, as n grows, the Binomial PMF **converges pointwise** to the Poisson PMF. For each fixed $k \geq 0$,

$$\mathbb{P}[X_n = k] \xrightarrow{n \uparrow \infty} \frac{e^{-\lambda} \lambda^k}{k!}.$$

This approximation is useful when there are many independent trials, each with a small success probability, and np is of moderate size.

Poisson approximation of Binomial



$\text{Bin}\left(100, \frac{4}{100}\right)$ vs. $\text{Pois}(4)$

Negative binomial

X has the **Negative Binomial** distribution with parameters $r \in \mathbb{N}$ and $p \in (0, 1)$ if

X = number of **failures** before r^{th} success
in independent $\text{Bern}(p)$ trials.

Write $X \sim \text{NegBin}(r, p)$. PMF:

$$\mathbb{P}[X = k] = \binom{k + r - 1}{r - 1} p^r (1 - p)^k, \quad \text{for } k = 0, 1, 2, \dots$$

Poisson and Negative Binomial are common models for count data.

Probability density functions

For a continuous X , its PDF f_X is the **continuous analog** of the PMF.

To calculate probabilities related to continuous r.v. X , we integrate f_X over the desired region. E.g.

$$\mathbb{P}[X \leq b] = \int_{-\infty}^b f_X(x) \, dx$$

$$\mathbb{P}[X \geq a] = \int_a^{\infty} f_X(x) \, dx$$

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f_X(x) \, dx$$

That is, for continuous r.v.'s,

Probability = Area under curve of PDF

Properties of PDF

A PDF must satisfy (similarly to PMF)

- ▶ **Non-negativity.** $f_X(x) \geq 0$
- ▶ **Normalization.** $\int_{-\infty}^{\infty} f_X(x) = 1$

Features about continuous r.v.'s and PDFs

- ▶ We have little occasion to ever write $\mathbb{P}[X = x]$ since

$$\mathbb{P}[X = x] = \int_x^x f_X(t) dt = 0 \quad \underline{\text{always}}$$

- ▶ Can be loose with endpoints since there is no ‘mass’ at any particular point:

$$\mathbb{P}[a \leq X \leq b] = \mathbb{P}[a \leq X < b] = \mathbb{P}[a < X < b]$$

Intuition about PDF

The **value** of the PDF f_X at a particular point x represents how *relatively* likely it is for X to be **near** x .

For a small width Δx ,

$$\mathbb{P}[x \leq X \leq x + \Delta x] \approx f_X(x) \Delta x.$$

- ▶ $f_X(x)$ is a **density**, not a probability.
- ▶ A density may be larger than 1; its total area must equal 1.
- ▶ Probabilities come from **areas over intervals**, not heights at points.

Gaussian (or Normal)

X has the **Gaussian** (or Normal) distribution with mean μ and variance σ^2 , denoted $X \sim \mathcal{N}(\mu, \sigma^2)$, if it has **PDF** supported on $\mathbb{R}$:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

When $X \sim \mathcal{N}(0, 1)$, it has the **standard Gaussian** (or standard Normal) distribution.

Linear functions of Gaussians are Gaussians:

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$.

Conversely, if $X \sim \mathcal{N}(0, 1)$, then $\mu + \sigma X \sim \mathcal{N}(\mu, \sigma^2)$.

CDF of Gaussian

Let $Z \sim \mathcal{N}(0, 1)$. Its **CDF** is often denoted

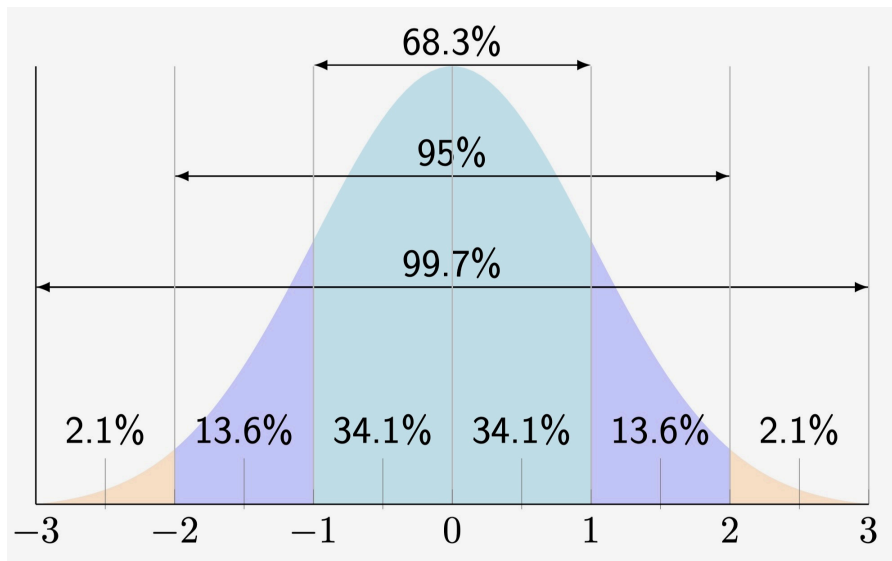
$$\Phi(x) = \mathbb{P}[Z \leq x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt.$$

This has **no closed form!** Precise numerical approximations are available.

Symmetry about zero For $x > 0$,

- ▶ $\mathbb{P}[Z \leq -x] = \Phi(-x) = \mathbb{P}[Z \geq x]$
- ▶ $\mathbb{P}[|Z| \leq x] = 2\Phi(x) - 1$

Gaussian statistics



For $Z \sim \mathcal{N}(0, 1)$,

$$\mathbb{P}[-1 < Z < 1] = \Phi(1) - \Phi(-1) \approx 0.683$$

$$\mathbb{P}[-2 < Z < 2] = \Phi(2) - \Phi(-2) \approx 0.95$$

$$\mathbb{P}[-3 < Z < 3] = \Phi(3) - \Phi(-3) \approx 0.997$$

Gamma

A **Gamma** random variable $X \sim \text{Gamma}(\alpha, \beta)$, with rate parameter β , takes positive real values. PDF:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad 0 < x < \infty$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ is the **gamma** function. This extends the factorial function to all positive reals, with $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$.

For particular choices of α and β , this (rather complicated) density specializes to many important distributions like **Exponential** or χ^2 . It provides a flexible model in statistics.

Chi-squared distribution

A chi-squared random variable $X \sim \chi^2(n)$ is a special case of a Gamma random variable, $\text{Gamma}(n/2, 1/2)$. PDF:

$$f_X(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, \quad \text{for } x > 0.$$

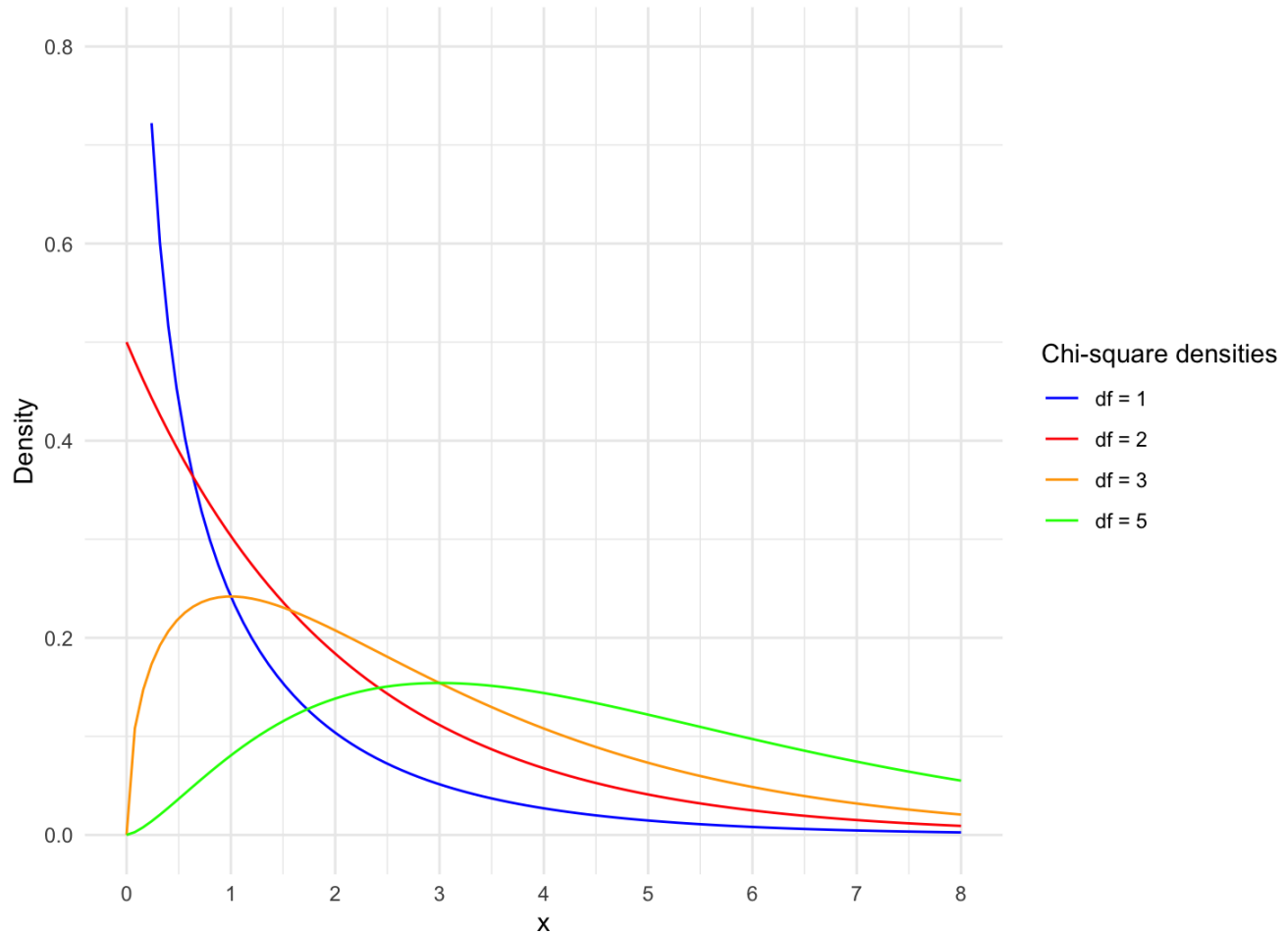
The parameter n is called the “degrees of freedom”.

Useful representation (often more so than PDF):

If $Z_1, \dots, Z_n \sim \mathcal{N}(0, 1)$ i.i.d., then

$$Z_1^2 + Z_2^2 + \dots + Z_n^2 \sim \chi^2(n)$$

Chi-squared distribution



Outline

Random variables

PMF

PDF

CDFs and quantiles

Expectation

Variance

Deviation inequalities

Cumulative distribution functions

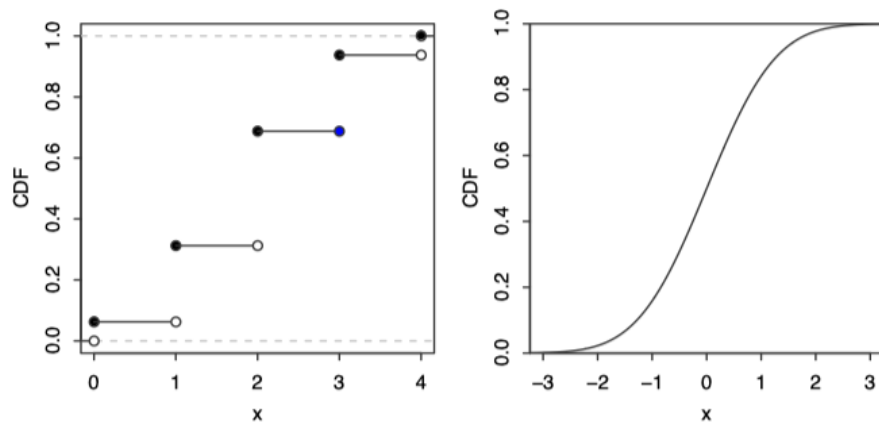
The **cumulative distribution function** (CDF) of a random variable X is

$$F_X(x) = \mathbb{P}[X \leq x]$$

This is a function from $\mathbb{R}$ to $[0, 1]$.

Properties:

- ▶ Non-decreasing
- ▶ Limits at infinities: $F_X(-\infty) = 0$ and $F_X(\infty) = 1$



What's the point of CDF?

CDF's **encode the distribution** of an r.v. E.g. for discrete r.v. X , they are an alternative to PMF's:

$$\mathbb{P}[X = 5] = \mathbb{P}[X \leq 5] - \mathbb{P}[X \leq 4] = F_X(5) - F_X(4).$$

However, CDF's are not particularly helpful for **discrete** r.v.'s, because of many discontinuities. Might as well study PMF.

CDF's will be useful when they are mostly *smooth* functions, which happens for **continuous** r.v.'s.

Important relation:

$$\frac{d}{dx} F_X(x) = \frac{d}{dx} \int_{-\infty}^x f_X(t) dt = f_X(x)$$

Takeaway: One may find the PDF by first finding the CDF and then differentiating.

Quantiles

The CDF $F(x)$ is non-decreasing

$$F(x) \leq F(y) \quad \text{for all } x \leq y$$

If $F : \mathbb{R} \rightarrow (0, 1)$ is continuous and *strictly* increasing, then it has a continuous inverse $F^{-1} : (0, 1) \rightarrow \mathbb{R}$.

F^{-1} is called the **quantile function** of X . For any $\alpha \in (0, 1)$, the α **quantile** (equivalently, the (100α) th **percentile**) is denoted

$$q_\alpha = F^{-1}(\alpha)$$

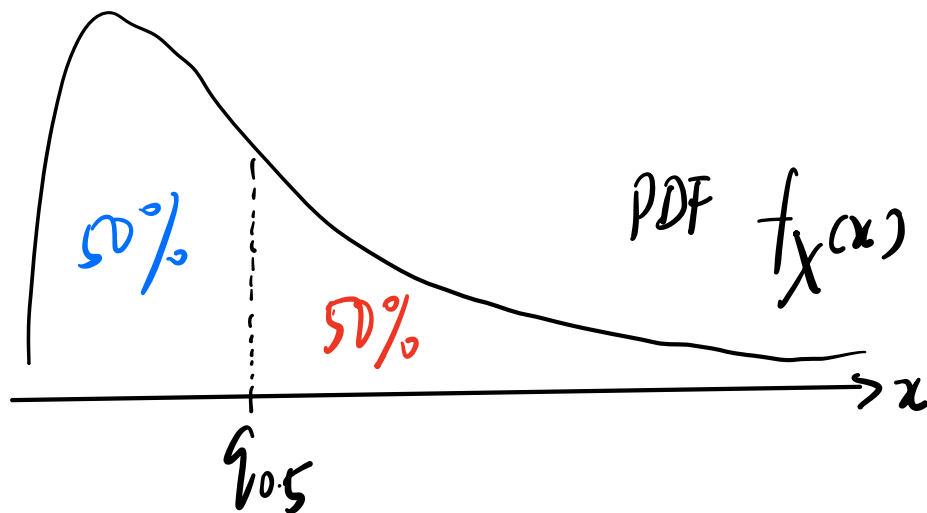


The number $q_{0.5}$ is the **median** of X .

Why is $q_{0.5}$ the median?

By definition $q_{0.5} = F^{-1}(0.5)$. So for a continuous, strictly increasing distribution,

$$\mathbb{P}[X \leq q_{0.5}] = F(q_{0.5}) = 0.5, \quad \text{and} \quad \mathbb{P}[X \geq q_{0.5}] = 0.5$$



Thus $q_{0.5}$ divides the probability mass into a **lower half** and an **upper half**.

Reconciling with discrete observations

Suppose we sort the observations:

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}.$$

We implicitly treat each observed occurrence as equally likely, with probability $1/n$. The **empirical CDF** records the fraction of observations at or below x :

$$\widehat{F}_n(x) = \frac{\#\{i : x_i \leq x\}}{n}.$$

Its median is therefore the location where half of the observations have accumulated.

- ▶ If $n = 2m + 1$, then the median is $x_{(m+1)}$: the middle-ranked observation.
- ▶ If $n = 2m$, every point in $[x_{(m)}, x_{(m+1)}]$ is a median in the balancing sense. The generalized inverse below chooses $x_{(m)}$; the usual convention reports

$$\frac{x_{(m)} + x_{(m+1)}}{2}.$$

Why do we need a generalized inverse?

The ordinary formula $q_\alpha = F^{-1}(\alpha)$ assumes that F is continuous and strictly increasing.

For a discrete or empirical distribution, the CDF is a **step function**:

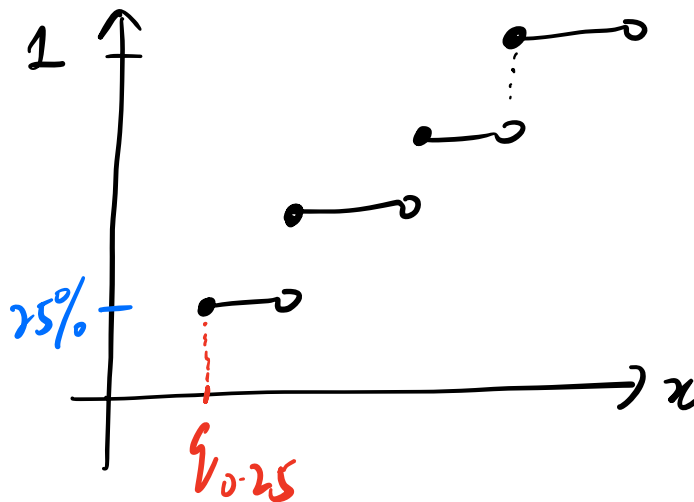
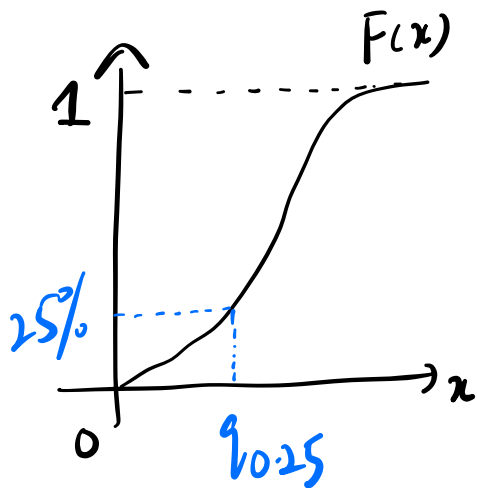
- ▶ A **jump** may skip the level α , so $F(x) = \alpha$ has no solution.
- ▶ A **flat piece** may make the inverse ambiguous.

Key idea: Monotonicity still lets us move from left to right and stop at the first location where the CDF *reaches or exceeds* α .

$$q_\alpha := F^{-1}(\alpha) = \inf \{x \in \mathbb{R} : F(x) \geq \alpha\}$$

Quantiles for general distributions

$$q_\alpha := F^{-1}(\alpha) = \inf \{x \in \mathbb{R} : F(x) \geq \alpha\}.$$



- Coincides with ordinary inverse if it exists.
- The same definition now works for continuous, discrete, and empirical distributions.

Interquartile range

Recall that $q_\alpha = F^{-1}(\alpha)$ denotes the α^{th} quantile of a distribution. Then

$$Q_1 = q_{0.25}, \quad M = q_{0.5}, \quad Q_3 = q_{0.75}$$

are the **first quartile**, **median**, and **third quartile**.

The **interquartile range** is

$$\boxed{\text{IQR} = Q_3 - Q_1}.$$

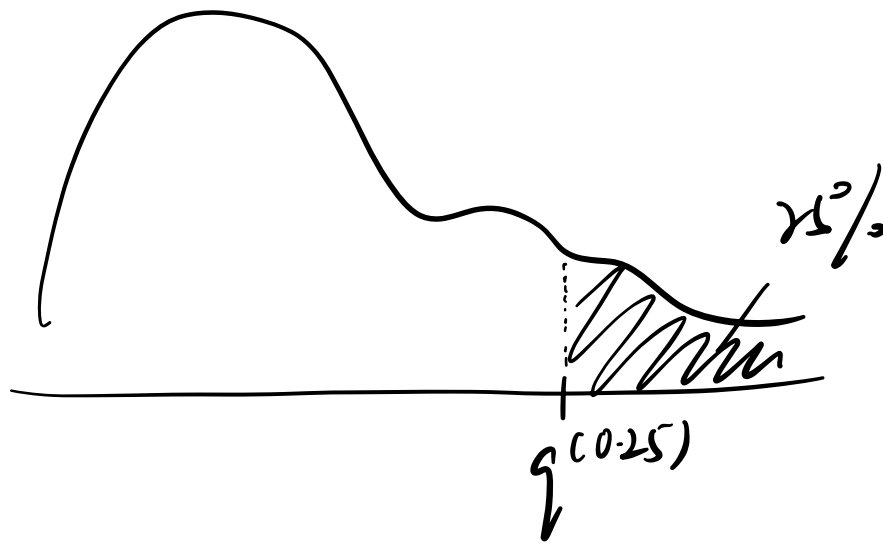
The interval $[Q_1, Q_3]$ contains (at least) 50% of the population distribution.

The median and IQR are less sensitive to extreme tails than the mean and standard deviation.

Upper-alpha point

Suppose F is continuous and strictly increasing. Define the **upper- α point** by

$$q^{(\alpha)} := q_{1-\alpha} = F^{-1}(1 - \alpha)$$



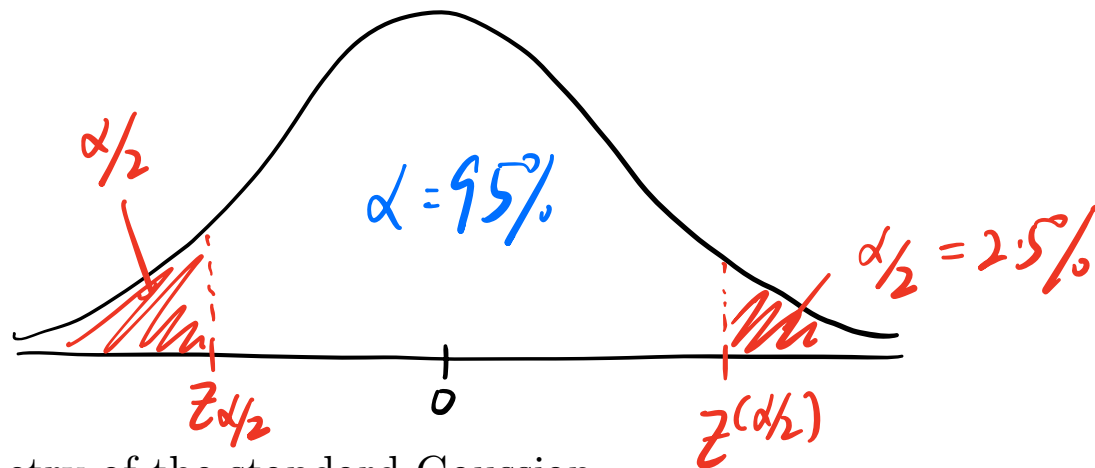
These cutoffs will appear repeatedly in **confidence intervals** and **hypothesis tests**.

Example: Standard Gaussian quantiles

Let $Z \sim \mathcal{N}(0, 1)$. We use the notation

$$z_\alpha = \Phi^{-1}(\alpha) \quad \alpha\text{-th quantile}$$

$$z^{(\alpha)} = \Phi^{-1}(1 - \alpha) \quad \text{upper-}\alpha \text{ point}$$



By symmetry of the standard Gaussian,

$$z_\alpha = -z^{(\alpha)}.$$

For instance, $z^{(0.025)} \approx 1.96$, and $z_{0.025} \approx -1.96$. Consequently,

$$\mathbb{P}[-1.96 \leq Z \leq 1.96] \approx 0.95$$

Universality of the Uniform

Suppose F is a continuous, strictly increasing CDF.

Theorem

1. If $U \sim \text{Unif}(0, 1)$ and $X = F^{-1}(U)$, then

$$\boxed{X \sim F}.$$

2. If $X \sim F$ and $U = F(X)$, then

$$\boxed{U \sim \text{Unif}(0, 1)}.$$

A particularly useful application is **simulation**: starting from the uniform U , we can generate random variables from a desired distribution F , provided F^{-1} is tractable to compute.

Universality of the Uniform

F: some cont., strictly ↑ CDF

Prove both parts of the theorem by:

1. If $U \sim \text{Unif}(0, 1)$, find the CDF of $F^{-1}(U)$.
2. If X has CDF F , find the CDF of $F(X)$.

$$\textcircled{1} \quad P[F^{-1}(U) \leq x] = P[U \leq F(x)] = F(x)$$

i.e. CDF of $F^{-1}(U)$ is F

$$\textcircled{2} \quad P[F(X) \leq u] = P[X \leq F^{-1}(u)]$$
$$= F(F^{-1}(u)) = u, \quad 0 < u < 1$$

Deduce $F(X) \sim \text{Unif}(0, 1)$

Example

The **Logistic distribution** has CDF

$$F(x) = \frac{e^x}{1 + e^x}, \quad x \in \mathbb{R}.$$

Let $U \sim \text{Unif}(0, 1)$.

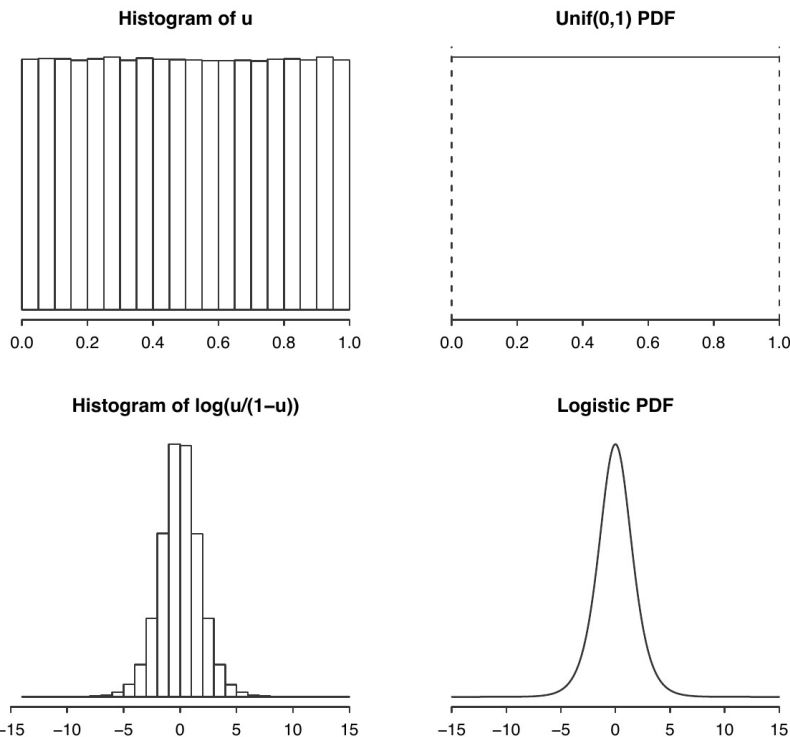
1. Find the quantile function $F^{-1}(u)$.
2. Use U to construct a random variable with the Logistic distribution.

$$u = \frac{e^x}{1+e^x} \Leftrightarrow e^x = \frac{u}{1-u} \Leftrightarrow F^{-1}(u) = \log \frac{u}{1-u}$$

$$\text{Then } X \sim F^{-1}(U) = \log\left(\frac{U}{1-U}\right) \sim \text{Logistic}$$

Universality with Logistic

$$U \sim \text{Unif}(0,1) \quad \Longrightarrow \quad F^{-1}(U) = \log\left(\frac{U}{1-U}\right) \sim \text{Logistic}.$$



Source: Blitzstein and Hwang, *Introduction to Probability*, 2nd ed., Figure 5.7.

Outline

Random variables

PMF

PDF

CDFs and quantiles

Expectation

Variance

Deviation inequalities

Expectation

The **expectation** (or **expected value**, or **mean**) of a discrete r.v. X is

$$\mathbb{E}[X] = \sum_{x \in S} x \cdot p_X(x)$$

Analogously for a continuous r.v. X ,

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

Expectation is a *weighted average*. It is weighted by the PMF or PDF.

Linearity of expectation

Linearity of expectation

For any r.v.'s X and Y , and for any constant c ,

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

$$\mathbb{E}[cX] = c\mathbb{E}[X]$$

Extends to any **finite linear combination**:

$$\mathbb{E}[c_1X_1 + \cdots + c_nX_n] = c_1\mathbb{E}[X_1] + \cdots + c_n\mathbb{E}[X_n].$$

This is an **extremely useful** property.

Example

Let $X \sim \text{Bin}(n, p)$. Then

$$\mathbb{E} X_i = p$$

$$\boxed{\mathbb{E}[X] = np}$$

Recall $X = \sum_{i=1}^n X_i$, $X_i \stackrel{\text{iid}}{\sim} \text{Bern}(p)$

$$\mathbb{E} X = \mathbb{E} \sum_{i=1}^n X_i = \sum_{i=1}^n \mathbb{E} X_i = np$$

(Easier than using Binomial PMF)

Law of the unconscious statistician (LOTUS)

Let X be an r.v., and g be a function. Then for discrete X ,

$$\mathbb{E}[g(X)] = \sum_{x \in S} g(x) \cdot p_X(x)$$

For continuous X ,

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

For instance

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) \, dx$$

In other words, we don't have to find the PMF/PDF of $g(X)$!

Note: This is **not** saying $\mathbb{E}[g(X)] = g(\mathbb{E}[X])$.

This is not true in general! (Unless g is linear).

Discrete vs. Continuous

PMF	PDF
$p_X(x) \geq 0$	$f_X(x) \geq 0$
$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x) \mathrm{d}x = 1$
$\mathbb{P}[X \in A] = \sum_{x \in A} p_X(x)$	$\mathbb{P}[X \in A] = \int_A f_X(x) \mathrm{d}x$
$\mathbb{E}[X] = \sum_x x \cdot p_X(x)$	$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) \mathrm{d}x$
$\mathbb{E}[g(X)] = \sum_x g(x) \cdot p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \mathrm{d}x$

Outline

Random variables

PMF

PDF

CDFs and quantiles

Expectation

Variance

Deviation inequalities

Variance, standard deviation

The variance of r.v. X is a number that measures the **spread** of a distribution.

$$\begin{aligned}\mathrm{Var}[X] &= \mathbb{E}[(X - \mathbb{E}X)^2] \\ &= \mathbb{E}[X^2] - (\mathbb{E}X)^2\end{aligned}$$

Variance is **constant shift invariant** and has **scaling property**:

$$\mathrm{Var}[aX + b] = a^2 \mathrm{Var}[X].$$

Standard deviation: interpretable on the scale of X rather than X^2 :

$$\mathrm{SD}[X] = \sqrt{\mathrm{Var}[X]}$$

Outline

Random variables

PMF

PDF

CDFs and quantiles

Expectation

Variance

Deviation inequalities

Motivation

Toss a coin 100 times.

Statistical questions of interest:

1. Suppose the coin was fair. Getting at least 75 heads is probably unlikely. **How unlikely?**
2. Suppose we end up with 75 heads. This indicates the coin is probably biased.
 - ▶ **How biased is it? I.e. what is p ?**
 - ▶ **How confident are we of our estimate p ?**

Deviation inequalities are useful tools that provide bounds on tail probabilities.

Markov's inequality

Let X be a **non-negative** random variable (i.e. $X \geq 0$ always).
Let $a > 0$. Then

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

Intuition: If the average salary is \$1000, then the fraction of people earning \$3000 cannot exceed $1/3$.

Takeaway: A non-negative r.v. cannot far exceed its expectation with high probability.

Example

Toss a fair coin 100 times. How unlikely is it to get at least 75 heads?

$$\text{Let } X \sim \text{Bin}(100, \frac{1}{2}). \quad \mathbb{E}X = 50.$$

$$\mathbb{P}[X \geq 75] \stackrel{\text{Markov}}{\leq} \frac{\mathbb{E}X}{75} = \frac{50}{75} = \frac{2}{3}$$

Useful, but probably too conservative.

Chebyshev's inequality

Let X be an **arbitrary** random variable and $d > 0$. Then

$$\mathbb{P}[|X - \mathbb{E}X| \geq d] \leq \frac{\text{Var}[X]}{d^2}$$

Chebyshev often leads to **better bounds** than Markov. But we require **more information**: the variance.

Actually follows from Markov's:

$$\begin{aligned} \mathbb{P}[|X - \mathbb{E}X| \geq d] &= \mathbb{P}[(X - \mathbb{E}X)^2 \geq d^2] \\ &\stackrel{\text{Markov}}{\leq} \frac{\mathbb{E}[(X - \mathbb{E}X)^2]}{d^2} \end{aligned}$$

non-neg. (pointing to the square term in the first equation)
Var X (pointing to the numerator in the second equation)

Example

(Revisited) Toss a fair coin 100 times. How unlikely is it to get at least 75 heads?

$$X \sim \text{Bin}(100, \frac{1}{2}), \quad \mathbb{E}X = 50, \quad \text{Var}X = 25.$$

$$\mathbb{P}[X \geq 75] = \mathbb{P}[X - 50 \geq 75 - 50] = \mathbb{P}[X - 50 \geq 25]$$

$$\underbrace{\{ |X - \mathbb{E}X| > d \}}_B = \underbrace{\{ X - \mathbb{E}X > d \}}_A \cup \{ X - \mathbb{E}X < -d \}$$
$$A \subseteq B \Rightarrow \mathbb{P}[A] \leq \mathbb{P}[B]$$

$$\leq \mathbb{P}[|X - 50| \geq 25] \stackrel{\text{Chebyshev's}}{\leq} \frac{25}{25^2} = 0.04$$

Better than Markov's.

Actual value 2.8×10^{-7}

Central limit theorem will give 2.9×10^{-7} (very accurate)

Chebyshev's inequality alternative form

Let X have standard deviation σ_X . Then

$$\mathbb{P}[|X - \mathbb{E}X| \geq C \cdot \sigma_X] \leq \frac{1}{C^2}$$

Interpretation: The probability that X is C standard deviations away from its mean $\mathbb{E}[X]$ is less than $1/C^2$.

