Covered topics

Math 431 - Lecture 2

Spring 2024

Disclaimer: although the plan is to have a fairly detailed list of the covered topics, the list below might not cover everything that we discussed in class.

Week 1

W 1/24: probability space (sample space, collection of events, probability measure)
what is an event?
Axioms of probability (properties of the probability measure)
Examples: coin flip, several coin flips
Direct product of sets

Fr 1/26: Experiments with equally likely outcomes
Sampling (with replacement with order, without replacement with/without order)

Reading materials:
Sections 1.1-1.3

Week 2

M 1/29: Sampling examples
Probability spaces with infinitely many outcomes
The probability of eventually getting tails in a sequence of coin flips is 1.
Decomposing an event as the disjoint union of simpler events: the probability of getting heads in an even number of coin flips
Uniformly chosen point from the interval [0,1].

W 1/31: $P(A^c) = 1 - P(A)$
If $B \subset A$ then $P(B) \leq P(A)$.
How to prove P(eventually we will get heads with a fair coin) $= 1$ using the monotonicity property of the probability measure

Fr 2/2: Inclusion-exclusion to compute the probability of union of events
Two events: $P(A \cup B) = P(A) + P(B) - P(AB)$
Inclusion-exclusion formula for three and n events
Inclusion-exclusion examples
Definition of a random variable

distribution of a random variable
Discrete random variable, probability mass function

Reading materials:
Sections 1.4-1.5

Week 3

M 2/5: Conditional probability of A given B (with $P(B) > 0$): $P(A|B) = \frac{P(AB)}{P(B)}$
$P(\cdot|B)$ is a probability measure for any fixed B with $P(B) > 0$.
The multiplication rule for conditional probabilities

$$P(AB) = P(A|B)P(B), \quad P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2A_1)\cdots P(A_n|A_{n-1}\cdots A_1)$$

W 2/7: Definition of a partition
If $B_1, \dots, B_n$ is a partition with $P(B_i) > 0$ then

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Bayes' formula:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

Similarly with a partition $B_1, \dots, B_n$.

Fr 2/9: Applications of Bayes' formula
Independence of two events A and B:

$$P(AB) = P(A)P(B)$$

If A and B are independent then the same is true for (i) A and B^c (ii) A^c and B (iii) A^c and B^c.

Reading materials:
Sections 2.1-2.3

Week 4

M 2/12: independence on n events, various characterizations
Independence of events constructed from independent events
E.g.: if A, B and C are independent then $A \cup B$ and C^c are independent.

Independence of the random variables $X_1, \dots, X_n$:
for any collection of sets $B_1, \dots, B_n \in \mathbb{R}$:

$$P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i)$$

Equivalent definition for the independence of discrete random variables $X_1, \dots, X_n$:

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

for any $x_1, \dots, x_n$.
Example: sampling with and without independence

W 2/14: The probability space of repeated independent trials with the same success probability.
Named distributions constructed from a sequence of independent trials with the same success probability

- Bernoulli with parameter p (the outcome of a single trial)
- Binomial with parameters n and p (the number of successes out of n trials)

Fr 2/16: Geometric distribution with parameter p (the number of trials needed for the first success)
The hypergeometric distribution.
Conditional independence of events.

Reading materials:
Sections 2.4-2.5

Week 5

M 2/19: The birthday problem: exact formula, and how to estimate it using the fact that $e^{-x} \approx 1 - x$ for small x.
Continuous distributions: the definition of the probability density function

Basic properties of the probability density function, how to identify a p.d.f.

How to compute probabilities using a p.d.f.

The uniform distribution on $[a, b]$

W 2/21: How the value of the probability density function can be connected to the probability of the random variable being in a small interval.

The cumulative distribution function of a random variable, definition, basic properties

How does the c.d.f. of a discrete and a continuous random variable look like.

Fr 2/23: How to identify the probability mass function from the c.d.f. and vice versa.

How to identify the c.d.f. of a continuous random variable, and how to compute its p.d.f. from the c.d.f.

The expected value of a random variable as the weighted average of possible values.

Computing the expected value for discrete and continuous random variables.

Computing the expected value of uniform, Bernoulli, and binomial random variables.

Expected value of an indicator random variable.

Reading materials:
Sections 3.1-3.3

Week 6

M 2/26: Computing the expected value of geometric random variable.

Random variables with infinite or undefined expectation

Expectation of a function of a random variable: computing $E[g(X)]$ for discrete and continuous X

The n^{th} moment of a random variable

W 2/28: The variance of a random variable: $\mathrm{Var}(X) = E\left[(X - EX)^2\right]$.

Another way to compute the variance: $\mathrm{Var}(X) = E[X^2] - (E[X])^2$.

Variance uniform random variables.

Variance of binomial and geometric random variables random variables.

How does the expectation and variance changes under a linear function:

$$E[aX + b] = aE[X] + b, \qquad \mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

Fr 3/1: The standard normal (or gaussian) distribution (the pdf φ and cdf Φ, expectation and variance)

The symmetry of the standard normal: $\Phi(-x) = 1 - \Phi(x)$.

How to compute probabilities involving a standard normal random variable using the table in the Appendix

The normal distribution with parameters μ and $\sigma^2 > 0$

If $Z \sim \mathcal{N}(0, 1)$ then $\sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$.

Reading materials:
Sections 3.3-3.5

Week 7

M 3/4: If $Z \sim \mathcal{N}(0,1)$ then $\sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$. If $X \sim \mathcal{N}(\mu, \sigma^2)$ then $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Expressing probabilities involving a general normal random variable in terms of Φ

The Central Limit Theorem for binomial random variables: if $0 < p < 1$ is fixed and $S_n \sim$Bin(n,p) then for any $a < b$ we have

$$\lim_{n\to\infty} P(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b) = \Phi(b) - \Phi(a)$$

In words: if we center a binomial random variable with its mean and divide it with its standard deviation then this scaled random variable will get closer and closer to a standard normal distribution as $n \to \infty$

Practical use: if $np(1-p) > 10$ and $S_n \sim$ Bin(n,p) then any probability involving $\frac{S_n-np}{\sqrt{np(1-p)}}$ can be approximated with the same probability involving a standard normal. E.g.

$$P(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b) \approx \Phi(b) - \Phi(a), \qquad P(\frac{S_n - np}{\sqrt{np(1-p)}} \le b) \approx \Phi(b)$$

W 3/6: Continuity correction: if $np(1-p) > 10$ and $S_n \sim$Bin(n,p) and we want to estimate a probability of the form of $P(k_1 \le S_n \le k_2)$ with k_1, k_2 integers, then it often helps to rewrite the probability as

$$P(k_1 \le S_n \le k_2) = P(k_1 - \frac{1}{2} \le S_n \le k_2 + \frac{1}{2})$$

and use the normal approximation for the modified endpoints.

The weak law of large numbers for binomial random variables.
Fix $0 < p < 1$ and $0 < \varepsilon$. Then if $S_n \sim$ Bin(n,p) then

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = 1.$$

Fr 3/8: Let $S_n \sim$Bin(n,p) denote the number of successes in a sequence of independent trials with an unknown success probability p. Then a natural estimate for p is $\hat{p}_n = \frac{S_n}{n}$ (the frequency of successes) and we can estimate the error using the normal approximation as

$$\begin{aligned} P(|\hat{p}_n - p| \le \varepsilon) &= P\left(-\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \le \frac{S_n - np}{\sqrt{np(1-p)}} \le \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \\ &\approx \Phi(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}) - \Phi(-\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}) = 2\Phi(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}) - 1 \\ &\ge 2\Phi(2\varepsilon\sqrt{n}) - 1 \end{aligned}$$

The definition of a confidence interval corresponding to a certain percentage.

Application to polling: for a large population sampling without replacement (which would give a hypergeometric distribution) is close to sampling with replacement (which gives a binomial distribution), and thus we can apply the normal approximation.

The Poisson(λ) distribution: probability mass function

If $\lambda > 0$ is fixed then the p.m.f. of Bin($n, \lambda/n$) converges to the p.m.f. of a Poisson(λ) distribution.

Reading materials:
Sections 4.1-4.4

Week 8

M 3/11: Mean and variance of the Poisson distribution

The Poisson approximation of the binomial distribution: if np^2 is not too big (say, less than 1) then a Bin(n, p) distribution can be approximated with a Poisson(np) random variable.

The Poisson distribution as a model for counting rare events.

The exponential distribution with parameter $\lambda > 0$: c.d.f., p.d.f.

The exponential distribution with parameter $\lambda > 0$ as the limit of $\frac{T_n}{n}$ where $T_n \sim$ Geo(λ/n). (waiting for a rare event in continuous time)

W 3/13: The expected value and variance of the exponential.

The memoryless property of the exponential.

The moment generating function of the random variable X: $M_X(t) = E(e^{tX})$.

Computing the moment generating function of various random variables

Computing the moments of a random variable using the moment generating function:

$$E(X^n) = \frac{d^n}{dt^n} M(t)\big|_{t=0}$$

if the moment generating function is finite in the neighborhood of 0.

Fr 3/15: Identifying the pmf of a discrete random variable from the moment generating function.

The moment generating function identifies the distribution of the random variable (if it is finite in a neighborhood of 0)

Computing the pmf of $g(X)$ if X is discrete

Computing the pdf of $Y = g(X)$ using the cdf method if X has a density function, g has a nonzero derivative apart from maybe finitely many points.

Outline:

- Write down the pdf and cdf of X (if possible)
- Identify the support of Y (if possible)
- Compute the cdf of Y by rewriting $P(Y \le y) = P(g(X) \le y)$ in terms of X and the cdf of X. (You will need to solve the inequality $g(X) \le y$ for X.)
- Differentiate the cdf of Y to get the pdf. (Be careful with case defined functions!)

Reading materials:
Sections 4.5, 5.1-5.2

Week 9

M 3/18: Additional examples for finding the pdf of $g(X)$ from g and the pdf of X.
Joint distribution of random variables
The joint probability mass function of random variables.
How to use the joint pmf to compute various probabilities about random variables. How to compute the expectation of a function of several discrete random variables
How to compute the marginal pmf from the joint pmf.

W 3/20: The multinomial distribution.
Jointly continuous random variables, the joint probability density function,
how to compute a probability involving jointly continuous random variables using the joint pdf
How to compute the expectation of a function of several jointly continuous random variables.
How to compute the marginal density function from the joint pdf.

Fr 3/22: The uniform distribution on 2 or 3 dimensional regions.
Zero probability events for jointly continuous random variables
Example for two continuous random variables that are not jointly continuous
The joint cumulative distribution function. Connection to the joint pdf for jointly continuous random variables.

Reading materials:
Sections 6.1-6.2, 6.4 (only the joint CDF)

Week 10

M 4/1: How to check for independence of discrete random variables form the joint pmf.
How to check for independence of jointly continuous random variables.
various examples,

W 4/3: (planned)
finding the minimum of two independent geometric random variables

Sum of two independent discrete or continuous random variables.
X, Y are independent and discrete then

$$P(X+Y=n) = \sum_k P(X=k)P(Y=n-k)$$

X, Y are independent and continuous then the pdf of $X+Y$ is given by

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$$

Sum of two independent binomials with the same success probability is also binomial

Fr 4/5: Sum of two independent geometric random variables with the same success probability: the negative binomial distribution.
Sum of two independent Poisson random variables is Poisson.
the sum of two independent standard normals
The sum of two independent exponentials with the same parameter.

Reading materials:
Sections 6.3, 7.1