Contents

Spring 2025, taught by Professor Kannan Ramchandran.

Chapter 1

Probability axioms

0.1 Probability space

A probability space: $(\Omega, \mathcal{F}, P)$:

- Ω: a set of all possible outcomes e.g. a binary string of length $2n$ with n ones
- $\mathcal{F}$: a set of all events (each event has 0 or more outcomes, $|\mathcal{F}| = 2^{|\Omega|}$)
- P: assignment of probability to event. Uniform sample space: $P(A) = \frac{|A|}{|\Omega|}$ e.g. $P(w) = \frac{1}{\binom{2n}{n}}$

Problem: There are n red and n blue balls.

Find $E[N]$, where N is the number of balls with the same color as the previous ball in the draw.

Solution: Let X_i be an indicator variable, whether balls i and $i-1$ share the same color, for $i = 2, \ldots, 2n$.

Then, we have: $\mathbb{E}[X_i] = \frac{n-1}{2n-1}$, since we "fix" a color of ith ball.

Thus, we have: $\mathbb{E}[N] = \mathbb{E}[\sum_{i=2}^{2n} X_i] = \sum_{i=2}^{2n} \mathbb{E}[X_i] = (2n - 2 + 1) \cdot \frac{n-1}{2n-1} = n - 1$.

0.2 Law of Total Probability

Given *disjoint* events A_i for $i = 1, \ldots, n$ that partition the sample space Ω:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i) = \sum_{i=1}^{n} \mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i)$$

Problem: Let $A_i = \{\text{exactly } i \text{ bins are empty}\}$.

Define $B = \{\text{all empty bins sit to the left of all bins containing at least one ball}\}$.

(a) Find $\mathbb{P}(B)$ in terms of A_i's.

(b) Calculate $\mathbb{P}(A_1)$.

Solution:

(a) Using the Law of Total Probability:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i) = \sum_{i=1}^{n} \mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i) = \sum_{i=1}^{n} \frac{1}{\binom{n}{i}} \mathbb{P}(A_i)$$

Note that the only relevant outcome for event A_1 is $0211\ldots11$, where ith value in the string is equal to the number of balls at bin i.

(b) $A_1 = \{\text{exactly one bin is empty}\}$, so it has to be some permutation of $\{0, 2, 1, \ldots, 1\}$.

- n ways to choose an empty bin
- $n-1$ ways to choose a bin with two balls
- $\binom{n}{2}$ to choose those two balls
- $(n-2)!$ ways to arrange (order matters) remaining $n-2$ balls into $n-2$ bins

$\mathbb{P}(A_1) = \frac{n(n-1)\binom{n}{2}(n-2)!}{n^n} = \frac{n!\binom{n}{2}}{n^n}$.

0.3 Union bound

Given random variables A_i for $i = 1, 2, \ldots, N$, the union bound is:

$\mathbb{P}(\bigcup_{i=1}^{N} A_i) \leq \sum_{i=1}^{N} \mathbb{P}(A_i)$

If events $A_1, \ldots, A_N$ are disjoint, $\sum_{i=1}^{N} \mathbb{P}(A_i) = \mathbb{P}(\bigcup_{i=1}^{N} A_i) \leq 1$.

A monkey types on a keyboard with 27 keys (corresponding to letters a-z, plus a period '.').

a) Assuming the monkey types each character independently and uniformly at random, what is the probability they type "class." on their first try? (Note the period at the end of the word 'class')

b) Let $\ell(i)$ denote the number of letters in the ith word in the English dictionary (all words are composed of only letters a-z, no periods). For example, if 'aardvark' is the 3rd word in the dictionary, then $\ell(3) = 8$. Assuming there are N words in the dictionary, use the axioms of probability to show that

$$\sum_{i=1}^{N} \left(\frac{1}{27}\right)^{\ell(i)} \leq 27.$$

Hint: Define N disjoint events.

a) Since we need to hit the 6 symbols in order, and keys are struck uniformly, the probability is $(1/27)^6$.

b) Let A_i be the probability the monkey types the ith word in the dictionary, followed by a period, on their first try. Then, $A_1, A_2, \ldots, A_N$ are disjoint events, and $P(A_i) = (1/27)^{\ell(i)+1}$. Hence, by the axioms of probability

$$\sum_{i=1}^{N} (1/27)^{\ell(i)+1} = \sum_{i=1}^{N} P(A_i) = P(\cup_{i=1}^{N} A_i) \leq 1.$$

Rearranging gives the claim.

Concepts tested: Computing simple probabilities from a given model; formulating suitable events; probability axioms.

Figure 1.1: `MT1 SP23 Q2`.

Chapter 2

Discrete and continuous RV

1 DRV

$X : \Omega \to \mathbb{R}$, e.g. $\Omega = \{1, 2, \ldots, 6\}$ for a dice.

$\mathbb{E}[X] = \sum_x x\,\mathbb{P}(X = x)$.

$\mathbb{E}[X|Y] = \sum_x x\,\mathbb{P}(X = x|Y)$.

1.1 Uniform

$X \sim \text{Unif}\{1, \ldots, n\}$.

$\mathbb{E}[X] = \sum_{i=1}^{n} i \cdot \frac{1}{n} = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$.

$\text{var}(X) = \frac{n^2-1}{12}$.

1.2 Bernoulli

$X \sim \text{B}(p)$, where $X = \begin{cases} 1 & \text{with probability} \quad p \\ 0 & \text{with probability} \quad 1-p \end{cases}$

$\mathbb{E}[X] = p$.

$\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p)$.

1.3 Binomial

$X \sim \text{Bin}(n, p)$ is n *independent* Bernoulli trials.

$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

$E[X] = \sum_{k=1}^{\infty} k \cdot \mathbb{P}_X(k) = \sum_{k=1}^{\infty} k \binom{n}{k} p^k (1-p)^{n-k}$.

Note: $\mathbb{E}[X] = \mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] = np$.

$\text{var}(X) = np(1-p)$.

Note: Bernoulli trials are independent, so the variance adds up.

1.4 Poisson

$X \sim \text{Poiss}(\lambda)$.

$\mathbb{P}(X=k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad x = 0, 1, 2, \ldots$

$M_X(s) = \sum_{k=0}^{\infty} e^{sk} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^s\lambda)^k}{k!} = e^{-\lambda} \cdot e^{\lambda e^s} = e^{-\lambda+\lambda e^s}$.

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

Note: $\mathbb{E}[X] = M_X'(s) = \lambda e^s e^{-\lambda+\lambda e^s}\big|_{s=0} = \lambda$.

$\text{var}(X) = \lambda$.

Note: $\mathbb{E}[X^2] = M_X''(s) = \lambda[e^s e^{-\lambda+\lambda e^s} + \lambda e^{2s} e^{-\lambda+\lambda e^s}]\big|_{s=0} = \lambda(1+\lambda)$.

Note: $\sum_{n=0}^{\infty} a^n = \frac{1}{1-a}$ and $\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$ is a Maclaurin expansion of Taylor series.

Poisson merging

Let $X \sim \text{Poiss}(\lambda), Y \sim \text{Poiss}(\mu)$ be independent RVs. Then $X + Y \sim \text{Poiss}(\lambda + \mu)$.

Poisson splitting

If $X \sim \text{Poiss}(\lambda)$, $Y|X = x \sim \text{Bin}(x, p)$, then $Y \sim \text{Poiss}(\lambda p)$.

Proof. $\mathbb{P}(Y = y) = \sum_x \mathbb{P}(Y = y|X = x) \cdot \mathbb{P}(X = x)$. □

1.5 Geometric

$X \sim \text{Geom}(p)$.

$\mathbb{P}(X = k) = p(1-p)^{k-1}$.

$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot p(1-p)^{k-1}$.

Note: Use Tail Sum Formula: $\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k) = \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{p}$.

Note: Use memoryless property:

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X|X=1] \cdot \mathbb{P}(X=1) + \mathbb{E}[X|X>1] \cdot \mathbb{P}(X>1) \\ &= 1 \cdot p + (1 + \mathbb{E}[X]) \cdot (1-p) = \frac{1}{p} \end{aligned} \tag{2.1}$$

$\text{var}(X) = \frac{1-p}{p^2}$.

Note: Given $g(X) = X^2$:

$$\begin{aligned} \mathbb{E}[g(X)] &= \mathbb{E}[g(X)|X=1] \cdot \mathbb{P}(X=1) + \mathbb{E}[g(X)|X>1] \cdot \mathbb{P}(X>1) \\ &= g(1) \cdot p + \mathbb{E}[g(X+1)] \cdot (1-p) = \frac{2-p}{p^2} \end{aligned} \tag{2.2}$$

Memoryless property of Geometric distribution

For integers $s > t > 0$, $\mathbb{P}(X > s|X > t) = \mathbb{P}(X > s - t)$ and $\mathbb{P}(X = s|X > t) = \mathbb{P}(X = s - t)$.

Proof. $\mathbb{P}(X > s|X > t) = \frac{\mathbb{P}[(X>s)\cap(X>t)]}{\mathbb{P}(X>t)} = \frac{\mathbb{P}(X>s)}{\mathbb{P}(X>t)} = \frac{(1-p)^s}{(1-p)^t} = (1-p)^{s-t}$. □

2 CRV

General formulas.

Leibniz integral rule: $\frac{d}{dx}\left(\int_{a(x)}^{b(x)} f(x,t)dt\right) = f(x,b(x)) \cdot \frac{d}{dx}b(x) - f(x,a(x)) \cdot \frac{d}{dx}a(x)$.

Fundamental Theorem of Calculus: $\boxed{\frac{d}{dx}\int_a^x f(t)dt = f(x)}$.

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

$\mathbb{P}(x \leq X \leq x+\epsilon) = \int_x^{x+\epsilon} f_X(t)dt \approx \epsilon f_X(x)$.

$F_X(x) = \int -\infty^\infty f_X(x)dx = \mathbb{P}(X \leq x)$.

$F_X(\infty) = 1$, $F_X(-\infty) = 0$.

$\frac{d}{dx}F_X(x) = f_X(x)$.

$f_X(x) = \int f_{X,Y}(x,y)dy$.

Conditional PDF: $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_X(x)f_{Y|X}(y|x)}{f_Y(y)}$.

2.1 Uniform

$X \sim \text{Unif}([a,b])$.

$f_X(x) = \frac{1}{b-a}$.

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < a \\ \frac{x-a}{b-a} & \text{if} \quad x \in [a,b] \\ 1 & \text{if} \quad x > b \end{cases}$$

$M_X(s) = \int_a^b e^{sx}\frac{1}{b-a}dx = \frac{e^{sb}-e^{sa}}{s(b-a)}$.

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

$\text{var}(X) = \frac{(b-a)^2}{12}$.

2.2 Exponential

$X \sim \text{Exp}(\lambda)$.

$f_X(x) = \lambda e^{-\lambda x}$, where $x \geq 0$.

$F_X(x) = \int_0^x \lambda e^{-\lambda\theta}d\theta = 1 - e^{-\lambda x}$.

$$M_X(s) = \mathbb{E}[e^{sX}] = \int_0^\infty \lambda e^{-\lambda x}e^{sx}dx = \lambda\int_0^\infty e^{(s-\lambda)x}dx = \begin{cases} \infty & \text{for} \quad s \geq \lambda \\ \frac{\lambda}{\lambda-s} & \text{for} \quad s < \lambda \end{cases}$$

$\mathbb{E}[X] = M'_X(s) = \frac{\lambda}{(\lambda-s)^2}\Big|_{s=0} = \frac{1}{\lambda}$.

$\mathbb{E}[X^2] = M''_X(s) = \frac{2\lambda}{(\lambda-s)^3}\Big|_{s=0} = \frac{2}{\lambda^2}$.

$\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$.

Note: Has a memoryless property like Geometric! $\mathbb{P}(X > s+t | X > s) = \mathbb{P}(X > t)$, where $0 < s < t$.

Problem: Exponential queue.

There are two cashiers. First one has a sevice time $X_1 \sim \text{Exp}(\lambda_1)$, second one $X_2 \sim \text{Exp}(\lambda_2)$.

Say you go to the first *available* cashier. There are two people ahead of you, at the first and seconds cashiers.

Suppose we know $\lambda_1 < \lambda_2$, i.e. the first cashier is slower.

(a) Find p_1, the probability the first cashier finishes before the second cashier.

(b) Find the probability that you are the last person to leave out of three.

(c) Find $\mathbb{E}[Y|Z]$, where $Y = \max(X_1, X_2)$, $Z = \min(X_1, X_2)$.

(d) Compute the joint density X_1 and $X_1 + X_2$.

Solution:

(a) Using Law of Iterated Expectation:

$$\begin{aligned}\mathbb{P}(X_1 < X_2) = \mathbb{E}[\mathbb{P}(X_1 < X_2|X_1)] = \mathbb{E}[e^{-\lambda_2 X_1}] = \int_0^\infty e^{-\lambda_2 x_1}\lambda e^{-\lambda_1 x_1} dx_1 \\ = \lambda \int_0^\infty e^{-x_1(\lambda_1+\lambda_2)} dx_1 = \frac{\lambda_1}{\lambda_1+\lambda_2}\end{aligned} \tag{2.3}$$

(b) Note that we want to minimize the wait time, and will choose the cashier that we believe will finish faster. Thus, we will choose the first cashier with probability p_1 and the second cashier with probability $p_2 = 1 - p_1$. We need to account for the cases when the chosen cashier is not the first one to finish:

$$p_1 \cdot p_2 + p_2 \cdot p_1 = 2p_1 p_2$$

(c) Due to memoryless property, we "take no breaks between trials" and the already passed time Z does not matter:

$$\mathbb{E}[Y|Z] = Z + p_1 \cdot \mathbb{E}[X_2] + p_2 \cdot \mathbb{E}[X_1] = Z + \frac{p_1}{\lambda_2} + \frac{p_2}{\lambda_1}$$

(d) Consider the joint CDF:

$$\begin{aligned}F_{X_1,X_1+X_2}(x,z) = \mathbb{P}(X_1 \le x, X_1 + X_2 \le z) &= \int_{-\infty}^{\infty} \mathbb{P}(X_1 \le x, X_1 + X_2 \le z | X_1 = x_1) f_{X_1}(x_1) dx_1 \\ &= \int_{-\infty}^{x} \mathbb{P}(X_2 \le z - x_1) f_{X_1}(x_1) dx_1 \\ &= \int_{-\infty}^{x} F_{X_2}(z - x_1) f_{X_1}(x_1) dx_1\end{aligned} \tag{2.4}$$

Differentiating both sides, we have:

$$f_{X_1,X_1+X_2}(x,z) = \frac{d}{dx_1}\frac{d}{dz}F_{X_1,X_1+X_2}(x,z) = \frac{d}{dx}\int_{-\infty}^{x} f_{X_2}(z-x_1)f_{X_1}(x_1)dx_1 = f_{X_2}(z-x)f_{X_1}(x) \tag{2.5}$$

Note that we are applying the Fundamental Theorem of Calculus here: $\boxed{\frac{d}{dx}\int_a^x f(t)dt = f(x)}$.

Problem: Let $X_i \sim \text{Exp}(\lambda)$ be independent for $1 \le i \le n$.
Let $M = \min(X_1, \ldots, X_n)$.

Solution: $\mathbb{P}(M > m) = \mathbb{P}(X_1 > m \cap \ldots X_n > m) = \mathbb{P}(X_1 > m) \cdot \cdots \cdot \mathbb{P}(X_n > m)$, since independent.
Then, we have: $\mathbb{P}(M > m) = e^{-n\lambda m}$.
Thus, we have: $F_M(m) = 1 - e^{-\lambda nm}$ and $M \sim \text{Exp}(\lambda n)$.

Problem: Let $X, Y \sim \text{U}([0,1])$.

Solution: Note that $\mathbb{P}(Y > X) = \mathbb{P}(Y < X) = \frac{1}{2}$ by symmetry.

$$\begin{aligned}\mathbb{E}[Y \,|\, \min(X,Y) = z] &= \mathbb{P}(Y \ge X) \cdot \mathbb{E}[Y | z \le Y \le 1] + \mathbb{P}(Y < X) \cdot \mathbb{E}[Y | Y = z] \\ &= \frac{1}{2}\frac{1}{1-z}\int_z^1 y dy + \frac{1}{2}z = \frac{1}{2}\frac{1}{1-z}\frac{1-z^2}{2} + \frac{1}{2}z = \frac{3z+1}{4}\end{aligned} \tag{2.6}$$

2.3 Gaussian

$X \sim \mathcal{N}(\mu, \sigma^2)$.
$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.
$M_X(s) = e^{\frac{\sigma^2 s^2}{2} + \mu s}$.

Standard Gaussian

$X \sim \mathcal{N}(0,1)$.
$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.
$\boxed{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\alpha x^2/2}dx = \frac{1}{\sqrt{\alpha}}.}$
$M_X(s) = \mathbb{E}[e^{sX}] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{sx}e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x^2-sx)}{2}}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{[(x-s)^2-s^2]}{2}}dx = e^{s^2/2}$.
$\boxed{\text{If } Y = \sigma X + \mu, \text{ i.e. } Y \sim \mathcal{N}(\mu,\sigma^2), \text{ then } M_Y(s) = e^{s\mu}M_X(\sigma s) = e^{\frac{\sigma^2 s^2}{2} + \mu s}.}$
$\mathbb{E}[X] = 0$.
$\text{var}(X) = 1$.

- $\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 = \text{var}(X) + \mathbb{E}[X]$.
- $\mathbb{E}[X^3] = M_X'''(s)\big|_{s=0} = 3$.

Problem: Consider $Z = X + Y$, where $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ are i.i.d.
Find $f_Z(z)$.

Solution: First, let's simplify: $\mu_X = \mu_Y = 0$, $\sigma_X^2 = \sigma_Y^2 = 1$.
$M_Z(z) = M_X(s) \cdot M_Y(s) = e^{s^2/2} \cdot e^{s^2/2} = e^{s^2}$.
Thus, $f_Z(z) \sim (0, 2)$.

Problem: Find $\mathbb{P}(X > Y)$, where $X \sim \mathcal{N}(0, 1)$, $Y \sim \mathcal{N}(2, 3)$ are independent.

Solution: Consider $Z = X - Y$. Note that $Z \sim \mathcal{N}(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2) = \mathcal{N}(-2, 4)$.

$$\mathbb{P}(X > Y) = \mathbb{P}(X - Y > 0) = \mathbb{P}(\mathcal{N}(-2, 4) > 0) = \mathbb{P}\left(\frac{Z+2}{\sqrt{4}} > \frac{0+2}{\sqrt{4}}\right) = \mathbb{P}(\mathcal{N}(0,1) \geq 1) = 1 - \Phi(1)$$

Jointly Gaussian independent RVs

Given independent $X \sim \mathcal{N}(\mu, \sigma_1^2)$ and $W \sim \mathcal{N}(0, \sigma_2^2)$. Consider $Y = X + W$, where $Y \sim \mathcal{N}(\mu, \sigma_1^2 + \sigma_2^2)$.
$X - \frac{\text{cov}(X,Y)}{\text{var}(Y)}(Y - \mu)$ is independent of Y.
In general, $X - \mathbb{E}[X|Y]$ is independent of Y.

3 General

3.1 Law of Iterated Expectation (LIE)

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

where $\mathbb{E}[X|Y] = g(Y)$ and is a RV.

Proof.

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|Y]] &= \sum_y \mathbb{E}[X|Y=y]\cdot\mathbb{P}(Y=y) = \sum_y\sum_x x\cdot\mathbb{P}(X=x|Y=y)\cdot\mathbb{P}(Y=y)\\ &= \sum_x x \sum_y \mathbb{P}(X=x,Y=y)\\ &= \sum_x x\cdot\mathbb{P}(X=x) = \mathbb{E}[X].\end{aligned} \tag{2.7}$$

□

A zoologist observes $B \sim \text{Poisson}(\mu)$ bears living on the prairie. Bear i has a territory with "range parameter" $R_i \sim \mathcal{N}(r, \sigma^2)$. Conditioned on R_i, the territory of bear i has area $X_i \sim \text{Uniform}(R_i^2, 3R_i^2)$, independent of the number of bears B. None of the bear territories overlap. What is the expected total territory area T occupied by all bears on the prairie?

This is like the "random sum of random variables" example we saw in class to illustrate the usefulness of iterated expectation. Since $T = \sum_{i=1}^{B} X_i$, and areas are independent of B, we can compute

$$\text{E}(T) = \text{E}(\text{E}(T \mid B)) = \text{E}(\text{E}(\textstyle\sum_{i=1}^{B} X_i \mid B)) = \text{E}(B\,\text{E}(X_1)) = \text{E}(B)\,\text{E}(X_1).$$

It remains to find the expected area of one bear's territory, which can also be evaluated by iterated expectation, since the distribution of X_i is uniform once we fix R_i:

$$\text{E}(X_1) = \text{E}(\text{E}(X_1 \mid R_1)) = \text{E}(2R_1^2) = 2(\text{var}(R_1) + \text{E}(R_1)^2) = 2(\sigma^2 + r^2).$$

So, the expected total area is

$$\text{E}(T) = 2\mu(\sigma^2 + r^2).$$

Concepts tested: Iterated expectation; linearity of expectation; variance decomposition in terms of second moment.

Figure 2.1: `MT1 SP23 Q3`.

3.2 Law of Total Variance (LTV)

$$\text{var}(X) = \mathbb{E}[\text{var}(X|Y)] + \text{var}(\mathbb{E}[X|Y])$$

Problem: Let $R \sim \text{Unif}\{1,2,3,4,5\}$ and $S \sim \mathcal{N}(R, R/2)$.

Solution: $\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S|R]]$.

Note that $E[S|R] = R$, since R is the mean of the Gaussian.

Then, we have: $\mathbb{E}[S] = \mathbb{E}[R] = \sum_r r\cdot\mathbb{P}(R=r) = \frac{1}{5}\sum_{i=1}^{5} i = \frac{1}{5}\frac{30}{2} = 3$.

Note that $\text{var}(S|R) = R/2$, since $R/2$ is the variance of the Gaussian.

$\text{var}(S) = \mathbb{E}[\text{var}(S|R)] + \text{var}(\mathbb{E}[S|R]) = \mathbb{E}[R/2] + \text{var}(R)$.

3.3 Tail Sum Formula

Let X be a RV that takes values only in $\mathbb{N}$. Then, $\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$.

Proof. $\mathbb{E}[X] = \sum_{k=1}^{\infty} k\,\mathbb{P}_X(k) = \sum_{k=1}^{\infty}(\sum_{l=1}^{k} 1)\,\mathbb{P}_X(k) = \sum_{l=1}^{\infty} \sum_{k=l}^{\infty} \mathbb{P}_X(X = k) = \sum_{l=1}^{\infty} \mathbb{P}(X \geq l)$. □

3.4 Variance

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

3.5 Variance of the sum of RVs

Let X, Y be RVs, then $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X, Y)$.

Corollary. If X, Y are independent, then $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.

Corollary. $\text{var}(2X) = 2\,\text{var}(X) + 2\,\text{cov}(X, X) = 4\,\text{var}(X)$.

3.6 Covariance and correlation

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

- If $\text{cov}(X, Y) = 0$, then X, Y are uncorrelated.
- If $\text{cov}(X, Y) > 0$, then when X increases, Y tends to increase.
- If $\text{cov}(X, Y) < 0$, then when X decreases, Y tends to decrease.

Properties:

1. $\text{cov}(X, Y) = \text{cov}(Y, X)$.
2. $\text{cov}(X, X) = \text{var}(X)$.
3. $\text{cov}(\alpha X + \beta, Y) = \alpha\,\text{cov}(X, Y)$.
4. $\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z)$.

Correlation coefficient: $\rho = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}} \in [-1, 1]$.

Problem: Given independent $X \sim \mathcal{N}(\mu, \sigma_1^2)$ and $W \sim \mathcal{N}(0, \sigma_2^2)$. Consider $Y = X + W$, where $Y \sim \mathcal{N}(\mu, \sigma_1^2 + \sigma_2^2)$.

Find $\text{cov}(X, Y)$.

Solution: $\text{cov}(X, X + W) = \text{cov}(X, X) + \text{cov}(X, W) = \text{var}(X) + 0 = \sigma_1^2$.

Problem: Toss a fair coin three times.

Define X = number of Heads in first 2 tosses and Y = number of Heads in all 3 tosses.

Solution: Note that $X \sim \text{Bin}(2, 1/2)$ and $Y \sim \text{Bin}(3, 1/2)$.

Let $Y = X + Z$, where $Z \sim \text{Bernoulli}(1/2)$.

Note that X and Z are independent, so $\text{cov}(X, Z) = 0$.

$\text{cov}(X, Y) = \text{cov}(X, X + Z) = \text{var}(X) + \text{cov}(X, Z) = 2\frac{1}{2}(1 - \frac{1}{2}) = \frac{1}{2} > 0$.

Intuitively, if there are more Heads in the first two tosses, there will be more Heads in all three tosses.

3.7 Dervided distributions

If a RV $Y = g(X)$ for some other RV X, then $\mathbb{E}[Y] = \mathbb{E}[g(X)]$.

3.8 Order statistics

Problem: Let $X_1, \ldots, X_n$ be i.i.d. RVs with common density $f_X(x)$ and CDF $F_X(x)$.

Let $X^{(k)}$ be the kth smallest of $(X_1, \ldots, X_n)$.

$X^{(1)}$ is the minimum, $X^{(n)}$ is the maximum.

What is the density $f_{X^{(k)}}(x)$ of $X^{(k)}$?

Solution: Note that $\mathbb{P}(X^{(k)} \in (x, x + dx)) \approx f_{X^{(k)}}(x)dx$.

In order for the kth smallest to lie on the interval $(x, x + dx)$:

- $k - 1$ points should lie on $(-\infty, x)$
- one point should lie on $(x, x + dx)$
- remaining $n - k$ points should lie on $(x + dx, \infty)$

Thus, we have:

$$f_{X^{(k)}}(x)dx \approx \mathbb{P}(X^{(k)} \in (x, x + dx)) = \binom{n-1}{k-1} F_X(x)^{k-1} \cdot n f_X(x)dx \cdot (1 - F_X(x))^{n-k} \tag{2.8}$$

Problem: Let variables be Uniform, so $X_i \sim \text{U}([0, 1])$, i.i.d. Find $\mathbb{E}[X^{(k)}]$ (see homework).

Solution: Note that $\boxed{\int_0^1 t^m(1-t)^n dt = \frac{m!n!}{(m+n+1)!}}$.

$$\begin{aligned}\mathbb{E}[X^{(k)}] &= \int_0^1 x f_{X^{(k)}}(x)dx = \int_0^1 x \cdot \binom{n-1}{k-1} F_X(x)^{k-1} \cdot n f_X(x) \cdot (1-F_X(x))^{n-k} dx \\ &= n\binom{n-1}{k-1}\int_0^1 x \cdot x^{k-1} \cdot 1 \cdot (1-x)^{n-k} dx = n\binom{n-1}{k-1}\int_0^1 x^k(1-x)^{n-k}dx \\ &= n\frac{(n-1)!}{(n-k)!(k-1)!}\frac{k!(n-k)!}{(n+1)!} = \frac{k}{n+1}\end{aligned} \tag{2.9}$$

Problem: Let variables be Exponential, so $X_i \sim \text{Exp}(\lambda)$, i.i.d. Find $\mathbb{E}[X^{(k)}]$.

Solution: Using the memoryless property, we can represent $X^{(k)} = Y_1 + \cdots + Y_k$, where Y_i is a waiting time between $X^{(i-1)}$ and $X^{(i)}$.

Note that Y_i is the minimum among remaining $n-i+1$ variables, and Y_i's are distributed with different rates λ_i.

$Y_1 = \min(X_1, \ldots, X_n) \sim \text{Exp}(\lambda n)$. Similarly, $Y_2 \sim \text{Exp}(\lambda(n-1))$, etc.

Then, we have:

$$\begin{aligned}\mathbb{E}[X^{(k)}] &= \mathbb{E}[Y_1 + \cdots + Y_n] = \sum_{i=1}^n \mathbb{E}[Y_i] \\ &= \sum_{i=1}^n \frac{1}{\lambda(n-i+1)} = \frac{1}{\lambda}\sum_{i=1}^n \frac{1}{n-i+1}\end{aligned} \tag{2.10}$$

3.9 Convolutions

Given $Z = X + Y$.

Note that $f_{Z|X}(z|x) = f_Y(z-x)$:

$$F_{Z|X}(z|x) = \mathbb{P}(Z \le z | X = x) = \mathbb{P}(X + Y \le z | X = x) = \mathbb{P}(x + Y \le z) = \mathbb{P}(Y \le z - x) = F_Y(z-x)$$

Then, we have a *convolution*:

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

3.10 Change of variables

See discussion 3 Q1.

Chapter 3

PMF, PDF, conditional PDF and MGF

PMF (Probability Mass Function) = PDF, but for discrete RV of the form $\mathbb{P}(X = k)$.

1 Conditional PDF

Problem: Consider $Y = \alpha X + Z$, where $X, Z \sim \mathcal{N}(0, 1)$ are i.i.d.

Find the consitional density of $X|Y$.

Solution: Then, $Y \sim \mathcal{N}(0, \alpha^2 + 1)$ and $Y|X = x \sim \mathcal{N}(\alpha x, 1)$. Note that $Y|x = \alpha x + Z$ with mean αx.

Using the Bayes rule $f_{X|Y} = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)}$, we have:

$$f_{X|Y} = \frac{1}{\sqrt{2\pi \frac{1}{\alpha^2+1}}} e^{\frac{-\left(x - \frac{\alpha}{\alpha^2+1} y\right)^2}{2 \cdot \frac{1}{\alpha^2+1}}} \tag{3.1}$$

i.e. $X|Y = y \sim \mathcal{N}(\frac{\alpha}{\alpha^2+1} y, \frac{1}{\alpha^2+1})$.

Then, we have: $\mathbb{E}[X|Y = y] = \frac{\alpha}{\alpha^2+1} y$ and $\mathbb{E}[Y|X = x] = \alpha x$.

2 MGF

Taylor expansion: $e^{sX} = 1 + sX + \frac{s^2 X^2}{2!} + \frac{s^3 X^3}{3!} + \dots$.

Here, X is a RV and s is a parameter:

$$M_X(s) = \mathbb{E}[e^{sX}] = 1 + s\,\mathbb{E}[X] + \frac{s^2}{2!}\,\mathbb{E}[X^2] + \frac{s^3}{3!}\,\mathbb{E}[X^3] + \dots$$

In general, $\frac{d^n}{ds^n}\,\mathbb{E}[e^{sX}]\big|_{s=0} = \mathbb{E}[X^n]$.

Moment-generating function of:

- a continous RV X: $M_X(s) = \mathbb{E}[e^{sX}] = \int_{-\infty}^{\infty} f_X(x) \cdot e^{sx} dx$.
- a discrete RV X: $M_X(s) = \mathbb{E}[e^{sX}] = \sum_k \mathbb{P}(X = k) \cdot e^{sx}$.

Problem: $M_X(s) = \frac{1}{2}e^{-3s} + \frac{1}{4}e^{2025s} + \frac{1}{4}e^{s}$.

Solution: Recognize discrete pattern: $M_X(s) = \sum_k e^{sk}\, \mathbb{P}(X = k)$.

Then, we have: $X = \begin{cases} -3 & \text{with probability } \frac{1}{2} \\ 2025 & \text{with probability } \frac{1}{4} \\ 1 & \text{with probability } \frac{1}{4} \end{cases}$

Properties:

1. $M_X(0) = 1$ (region of convergence).
2. If $Y = \alpha X + \beta$, $M_Y(s) = \mathbb{E}[e^{sY}] = \mathbb{E}[e^{s(\alpha X+\beta)}] = \mathbb{E}[e^{s\alpha X} \cdot e^{s\beta}] = e^{s\beta}\, \mathbb{E}[e^{s\alpha X}] = e^{s\beta} M_X(\alpha s)$.
3. If $Z = X + Y$, X, Y are independent RVs, then:

$$M_Z(s) = \mathbb{E}[e^{sZ}] = \mathbb{E}[e^{s(X+Y)}] = \mathbb{E}[e^{sX} \cdot e^{sY}] = M_X(s) \cdot M_Y(s)$$

   In general, if $Z = \sum_{i=1}^{n} X_i$, when X_i's are independent, we have:

$$M_Z(s) = \prod_{i=1}^{n} M_{X_i}(s)$$

4. $\boxed{\text{Joint MGF: } M_{X_1,X_2,\ldots,X_n}(s_1, s_2, \ldots, s_n) = \mathbb{E}[e^{s_1X_1+s_2X_2+\cdots+s_nX_n}] = \prod_{i=1}^{n} M_{X_i}(s_i).}$

   For example, for independent X, Y, we can show independence of $X - Y$, $X + Y$:
   $M_{X-Y,X+Y}(s_1, s_2) = \mathbb{E}[e^{s_1(X-Y)+s_2(X+Y)}] = \mathbb{E}[e^{(s_1+s_2)X+(s_2-s_1)Y}] = \mathbb{E}[e^{(s_1+s_2)X}] \cdot \mathbb{E}[e^{(s_2-s_1)Y}]$.

Note: MGF is always unique and strictly positive!

Problem: Consider $Z = X^2 + Y^2$, where $X, Y \sim \mathcal{N}(0, 1)$, i.i.d.
Find the density of Z.

Solution: Use MGF!
$M_{X^2+Y^2}(s) = M_{X^2}(s) M_{Y^2}(s)$.
We have:

$$M_{X^2}(s) = \mathbb{E}[e^{sX^2}] = \int_{-\infty}^{\infty} e^{sx^2} f_X(x) dx = \int_{-\infty}^{\infty} e^{sx^2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(1-2s)x^2/2} dx = \frac{1}{\sqrt{1-2s}}$$

Thus:

$$M_{X^2+Y^2}(s) = \frac{1}{1-2s} = \frac{1/2}{1/2 - s}$$

and $Z \sim \text{Exp}(1/2)$. Therefore, PDf of Z is $\lambda e^{-\lambda z} = \frac{1}{2}e^{-z/2}$.

Problem: Consider a coin that turns Heads with probabiltiy p. For a given integer $k \geq 1$, let N denote the number of independent flips until we see exactly k Heads.

Find PMF of N, $M_N(s)$ and $\text{var}(N)$.

Solution: Note that N is a DRV.

For the number of flips to be $N = n$, we need the nth toos to be Heads, and the previous $n-1$ tosses to have exactly $k-1$ Heads.

The PMF of N is:

$$\mathbb{P}(N = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

Let $N = X_1 + \cdots + X_k$, where X_i is a number of flips is takes for ith Heads to show up, i.e. $X_i \sim \text{Geom}(p)$.

Then, we have:

$$M_N(s) = \prod_{i=1}^{k} M_{X_i}(s)$$

where $M_{X_i}(s) = \mathbb{E}[E^{sX_i}] = \sum_x \mathbb{P}(X = x) \cdot e^{sx} = \sum_{x=1}^{\infty} p(1-p)^{x-1} e^{sx} = pe^s \sum_{x=0}^{\infty} [e^s(1-p)]^x = \frac{pe^s}{1-e^s(1-p)}$.

Thus, we have:

$$M_N(s) = \left(\frac{pe^s}{1-e^s(1-p)}\right)^k$$

Note that $\text{var}(N) = \text{var}(X_1 + \cdots + X_k) = \sum_{i=1}^{k} \text{var}(X_i)$, since X_i's are independent.

Therefore:

$$\text{var}(N) = k \cdot \frac{1-p}{p^2}$$

Chapter 4

Concentration inequalities

Limit behavior of RVs. We observe a sequence of i.i.d. RVs: $X_1, \ldots, X_n \sim X$.

Let $M_n = \frac{\sum_{i=1}^{n} X_i}{n}$ be the sample mean.

1. $\mathbb{E}[M_n] = \mathbb{E}[X]$, i.e. an unbiased estimate.
2. $\text{var}(M_n) = \frac{\text{var}(X)}{n}$, assuming $\text{var}(X) < \infty$.

As $n \to \infty$, $\mathbb{E}[M_n] = \mathbb{E}[X]$, but $\text{var}(M_n) \to 0$, i.e. starts to be more *deterministic*.

Tail bounds: upper bound on probability that a RV deviates from its mean or central value.

What happens to the "deviation" $|M_n - \mathbb{E}[X]|$ as n gets large?

Concentration bounds: how tightly a RV concetrates around its mean or central value.

How fast does $\text{var}(M_n) \to 0$?

0.1 Markov's inequality

If a random variable $X \geq 0$, then $\mathbb{P}(X \geq a) \leq \frac{E[X]}{a}$ for a constant $a > 0$.

Proof. $1_{X \geq a} \leq \frac{x}{a}$. Take $\mathbb{E}[\cdot]$: $\mathbb{E}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$. □

0.2 Chebyshev's bound

$$\mathbb{P}(|X - \mathbb{E}[X]| > c) \leq \frac{\text{var}(X)}{c^2}$$

Proof. $\mathbb{P}(|X - \mathbb{E}[X]| \geq c) = \mathbb{P}\left((X - \mathbb{E}[X])^2 \geq c^2\right) \leq \frac{\mathbb{E}[(X-\mathbb{E}[X])^2]}{c^2} = \frac{\text{var}(X)}{c^2}$. □

0.3 Chernoff's bound

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{sX} \geq e^{sa}) \leq \frac{\mathbb{E}[e^{sX}]}{e^{sa}} = \frac{M_X(s)}{e^{sa}} \quad \forall s > 0$$

Solve for the smallest RHS, so the tightest bound!

- $\mathbb{P}(X \geq a) \leq \inf_{s>0} \frac{M_X(s)}{e^{sa}}$.
- $\boxed{\mathbb{P}(X \leq a) \leq \inf_{s<0} \frac{M_X(s)}{e^{sa}}.}$

Problem: Let $X \sim \text{Bin}(n, p)$. Upper bound $\mathbb{P}(X \geq \alpha n)$, where $p < \alpha < 1$.

Solution: $X = Y_1 + \cdots + Y_n$, where $Y_i \sim \text{Bernoulli}(p)$ are independent.

Note that $M_{Y_i}(s) = \mathbb{E}[e^{sY_i}] = e^s \cdot p + e^0 \cdot (1-p) = pe^s + 1 - p$.

$M_X(s) = \prod_{i=1}^n M_{Y_i}(s) = (M_{Y_i}(s))^n = (pe^s + 1 - p)^n$.

Then, we have:

$$\mathbb{P}(X \geq \alpha n) = \inf_{s>0} \frac{M_X(s)}{e^{s\alpha n}} = \inf_{s>0} e^{-\alpha n s}(pe^s + 1 - p)^n \tag{4.1}$$

$-\alpha n e^{-\alpha n s}(pe^s + 1 - p)^n + e^{-\alpha n s} n (pe^s + 1 - p)^{n-1} pe^s = 0$.

Thus, it is minimized at $s = \ln\left(\frac{\alpha(1-p)}{p(1-\alpha)}\right)$.

Chapter 5

Convergence

1 WLLN, SLLN

- WLLN: every function of samples goes to mean.
- SLLN: every realization of samples goes to mean.

Convergence in probability. Given $X_1, \ldots, X_n \xrightarrow{p} X$,

$$\lim_{n \to \infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0 \quad \forall \epsilon$$

Weak Law of Large Numbers. If $X_1, \ldots, X_n \sim X$ are i.i.d. RVs with mean $\mathbb{E}[X] = \mu$ and finite variance,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \to 0 \text{ as } n \to \infty$$

i.e. $\forall \epsilon, \sigma > 0 : \ \exists N(\epsilon, \sigma)$ s.t. $\mathbb{P}(|M_n - \mu| \geq \epsilon) < \sigma \quad \forall n > N(\epsilon, \sigma)$.

Note: $M_n \xrightarrow{p} \mathbb{E}[X]$.

Remark: ϵ captures "accuracy level", σ captures "confidence level".

Proof. Note that $\mathbb{E}[M_n] = \mu$, the true mean.

Let $\text{var}(X_i) = \sigma_X^2 < \infty$, so $\text{var}(M_n) = (\frac{1}{n})^2 n \,\text{var}(X) = \frac{\sigma_X^2}{n}$.

Apply Chebyshev: $\mathbb{P}(|M_n - \mu| \geq \epsilon) \leq \frac{\text{var}(M_n)}{\epsilon^2} = \frac{\sigma_X^2}{n\epsilon^2} \to 0$ as $n \to \infty$.

We say $M_n \xrightarrow{p} \mu$ ("M_n converges to μ in probability"). □

Strong Law of Large Numbers.

2 CLT

Central Limit Theorem.

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq x) = \Phi(x) \quad \forall x$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$.

(Blank.)