Section 9

Review

Law of total probability and law of total expectation

1) **Law of Total Probability (partition based on value of a r.v.):** If X is a discrete random variable, then

$$\mathbb{P}(A) = \sum_{x \in \Omega_X} \mathbb{P}(A|X = x)p_X(x)$$

If X is a continuous random variable, then

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A|X = x)f_X(x)\ dx$$

2) **Conditional Expectation**: Let X and Y be random variables. Then, the conditional expectation of X given $Y = y$ is

$$\mathbb{E}[X|Y = y] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|Y = y) \qquad X \text{ discrete}$$

and for any event A,

$$\mathbb{E}[X|A] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x|A) \qquad X \text{ discrete}$$

Note that linearity of expectation still applies to conditional expectation: $\mathbb{E}[X + Y|A] = \mathbb{E}[X|A] + \mathbb{E}[Y|A]$

3) **Law of Total Expectation (Event Version)**: Let X be a random variable, and let events $A_1, ..., A_n$ partition the sample space. Then,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X|A_i]\mathbb{P}(A_i)$$

4) **Law of Total Expectation (RV Version)**: Suppose X and Y are random variables. Then,

$$\mathbb{E}[X] = \sum_{y} \mathbb{E}[X|Y = y]p_Y(y) \qquad Y \text{ discrete r.v..}$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y]f_Y(y)dy \qquad Y \text{ continuous r.v.}$$

Maximum Likelihood Estimation

1) **Realization/Sample**: A realization/sample x of a random variable X is the value that is actually observed.

2) **Likelihood**: Let $x_1, \ldots x_n$ be iid realizations from probability mass function $p_X(\mathrm{x}\ ; \theta)$ (if X discrete) or density $f_X(\mathrm{x}\ ; \theta)$ (if X continuous), where θ is a parameter (or a vector of parameters). We define the likelihood function to be the probability of seeing the data.

If X is discrete:

$$L(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} p_X(x_i \mid \theta)$$

If X is continuous:

$$L(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f_X(x_i \mid \theta)$$

3) Maximum Likelihood Estimator (MLE): We denote the MLE of θ as $\hat{\theta}_{\text{MLE}}$ or simply $\hat{\theta}$, the parameter (or vector of parameters) that maximizes the likelihood function (probability of seeing the data).

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} L(x_1, \ldots, x_n \mid \theta) = \arg\max_{\theta} \ln L(x_1, \ldots, x_n \mid \theta)$$

4) Log-Likelihood: We define the log-likelihood as the natural logarithm of the likelihood function. Since the logarithm is a strictly increasing function, the value of θ that maximizes the likelihood will be exactly the same as the value that maximizes the log-likelihood.

If X is discrete:

$$\ln L(x_1, \ldots, x_n \mid \theta) = \sum_{i=1}^{n} \ln p_X(x_i \mid \theta)$$

If X is continuous:

$$\ln L(x_1, \ldots, x_n \mid \theta) = \sum_{i=1}^{n} \ln f_X(x_i \mid \theta)$$

5) Steps to find the maximum likelihood estimator, $\hat{\theta}$:

(a) Find the likelihood and log-likelihood of the data.

(b) Take the derivative of the log-likelihood and set it to 0 to find a candidate for the MLE, $\hat{\theta}$.

(c) Take the second derivative and show that $\hat{\theta}$ indeed is a maximizer, that $\frac{\partial^2 L}{\partial \theta^2} < 0$ at $\hat{\theta}$. Also ensure that it is the global maximizer: check points of non-differentiability and boundary values.

(d) If we are finding the MLE for a set of parameters, then we set up the system of equations obtained by taking the partial derivative of the log-likelihood function with respect to each of the parameters and setting it equal to 0. We then solve this system to get the MLEs. (And again, second order conditions need to be checked.)

6) An estimator $\hat{\theta}$ for a parameter θ of a probability distribution is **unbiased** iff $\mathbb{E}[\hat{\theta}(X_1, \ldots, X_n)] = \theta$

Task 1 – Trapped Miner

A miner is trapped in a mine containing 3 doors.

- D_1: The 1^{st} door leads to a tunnel that will take him to safety after 3 hours.
- D_2: The 2^{nd} door leads to a tunnel that returns him to the mine after 5 hours.
- D_3: The 3^{rd} door leads to a tunnel that returns him to the mine after a number of hours that is Binomial with parameters $(12, \frac{1}{3})$.

At all times, he is equally likely to choose any one of the doors. What is the expected number of hours for this miner to reach safety? Use the law of total expectation.

Task 2 – Lemonade Stand

Suppose I run a lemonade stand, which costs me \$100 a day to operate. I sell a drink of lemonade for \$20. Every person who walks by my stand either buys a drink or doesn't (no one buys more than one). If it is raining, n_1 people walk by my stand, and each buys a drink independently with probability p_1. If it isn't raining, n_2 people walk by my stand, and each buys a drink independently with probability p_2. It rains each day with probability p_3, independently of every other day. Let X be my profit over the next week. In terms of n_1, n_2, p_1, p_2 and p_3, what is $\mathbb{E}[X]$? Use the law of total expectation.

Task 3 – Mystery Dish!

A fancy new restaurant has opened up that features only 4 dishes. The unique feature of dining here is that they will serve you any of the four dishes randomly according to the following probability distribution: give dish A with probability 0.5, dish B with probability θ, dish C with probability 2θ, and dish D with probability $0.5 - 3\theta$. Each diner is served a dish independently. Let x_A be the number of people who received dish A, x_B the number of people who received dish B, etc, where $x_A + x_B + x_C + x_D = n$. Find the MLE $\hat{\theta}$ for θ.

Task 4 – A Red Poisson

Suppose that $x_1, \ldots, x_n$ are i.i.d. samples from a Poisson(θ) random variable, where θ is unknown. In other words, they follow the distributions $\mathbb{P}(k; \theta) = \theta^k e^{-\theta}/k!$, where $k \in \mathbb{N}$ and $\theta > 0$ is a positive real number.

Find the MLE of θ.

Task 5 – A biased estimator

In class, we showed that the maximum likelihood estimate of the variance θ_2 of a normal distribution (when both the true mean μ and true variance σ^2 are unknown) is what's called the *population variance*. That is

$$\hat{\theta}_2 = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\theta}_1)^2)\right)$$

where $\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the MLE of the mean. Is $\hat{\theta}_2$ unbiased?

Task 6 – Weather Forecast

A weather forecaster predicts sun with probability θ_1, clouds with probability $\theta_2 - \theta_1$, rain with probability $\frac{1}{2}$ and snow with probability $\frac{1}{2} - \theta_2$. This year, there have been 55 sunny days, 100 cloudy days, 160 rainy days and 50 snowy days. What is the maximum likelihood estimator for θ_1 and θ_2?

Task 7 – Faulty Machines

You are trying to use a machine that only works on some days. If on a given day, the machine is working it will break down the next day with probability $0 < b < 1$, and works on the next day with probability $1 - b$. If it is not working on a given day, it will work on the next day with probability $0 < r < 1$ and not work the next day with probability $1 - r$.

a) In this problem we will formulate this process as a Markov chain. First, let $X^{(t)}$ be a variable that denotes the state of the machine at time t. Then, define a state space $\mathcal{S}$ that includes all the possible states that the machine can be in. Lastly, for all $A, B \in \mathcal{S}$ find $\mathbb{P}(X^{(t+1)} = A \mid X^{(t)} = B)$ (A and B can be the same state).

b) Suppose that on day 1, the machine is working. What is the probability that it is working on day 3?

c) As $n \to \infty$, what does the probability that the machine is working on day n converge to? To get the answer, solve for the *stationary distribution*.

Task 8 – Another Markov Chain

Suppose that the following is the transition probability matrix for a 4 state Markov chain (states 1,2,3,4).

$$M = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/3 & 0 & 0 & 2/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/5 & 2/5 & 2/5 & 0 \end{bmatrix}$$

a) What is the probability that $X^{(2)} = 4$ given that $X^{(0)} = 4$?

b) Write down the system of equations that the stationary distribution must satisfy and solve them.

Task 9 – Three Tails

You flip a fair coin until you see three tails in a row. Model this as a Markov chain with the following states:

- S: start state, which we are only in before flipping any coins.
- H: We see a heads, which means no streak of tails currently exists.
- T: We've seen exactly one tail in a row so far.

- TT: We've seen exactly two tails in a row so far.

- TTT: We've accomplished our goal of seeing three tails in a row, stop flipping, and stay there.

a) Write down the transition probability matrix.

b) Write down the system of equations whose variables are $D(s)$ for each state $s \in \{S, H, T, TT, TTT\}$, where $D(s)$ is the expected number of steps until state TTT is reached starting from state s. Solve this system of equations to find $D(S)$.

c) Write down the system of equations whose variables are $\gamma(s)$ for each state $s \in \{S, H, T, TT, TTT\}$, where $\gamma(s)$ is the expected number of heads seen before state TTT is reached. Solve this system to find $\gamma(S)$, the expected number of heads seen overall until getting three tails in a row.

Task 10 – Use of Law of Total Probability

Suppose that the time until server 1 crashes is $X \sim \text{Exp}(\lambda)$ and the time until server 2 crashes is independent, with $Y \sim \text{Exp}(\mu)$.
What is the probability that server 1 crashes before server 2?

Task 11 – Elevator rides

The number X of people who enter an elevator on the ground floor is a Poisson random variable with mean 10. If there are N floors above the ground floor, and if each person is equally likely to get off at any one of the N floors, independently of where others get off, compute the expected number of stops the elevator will make before discharging all the passengers.