CSE 312 Midterm – Taylor Blau

1 Combinatorics

1.1 Sum & Product Rule

The product rule states that for distinct events $A_1, \ldots, A_n$ each having $m_1, \ldots, m_n$ outcomes overall, the total number of outcomes is given as: $\prod_i m_i$.
The sum rule states that for the same scenario, where $A_1, \ldots, A_n$ are independent, that the total number of outcomes is instead given as: $\sum_i m_i$.

1.2 Ordering

The number of ways to order n distinct objects is $n!$. The number of ways to *permute* n objects (arrange a k-subset where order matters) is given as:

$$P(n,k) = \frac{n!}{(n-k)!}$$

Similarly, the number of ways to *combine* n objects (form a k-subset where order does not matter) is:

$$C(n,k) = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

1.3 Multinomial Coefficient

The number of ways to distinctly arrange n objects where $k < n$ of them are distinct is given as (where n_i represents the number of occurrences of the i-th distinct element) :

$$\frac{n!}{\prod_i n_i!} = \binom{n}{n_1, \ldots, n_k}$$

1.4 Pigeonhole Principle

If there are n pigeons and k holes where $n > k$, then there exists a hole with at least $\lceil n/k \rceil$ pigeons.

1.5 Binomial Theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k},\ x, y \in \mathbb{R}, n \in \mathbb{N}$$

1.6 Principle of Inclusion-Exclusion

Suppose A and B are non-distinct set (i.e., $A \cap B \neq \emptyset$), then:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

And so on for more sets, alternating adding and subtracting different size pairs.

1.7 Stars and Bars

For r bins and n objects:

Distinct objects, distinct bins r^n.

k-distinct objects, distinct bins $\binom{n}{k} r^k$.

Identical objects, distinct bins $\binom{n+r-1}{r-1}$.

Identical objects, non-empty distinct bins $\binom{n-1}{r-1}$.

2 Probability

2.1 Complementation

When an event E is difficult to count, but its complement $\overline{E}$ is easier, the following applies:

$$\mathbb{P}[E] = 1 - \mathbb{P}[\overline{E}]$$

This is useful in solving statements like "contains at least n", where the following is applied:

$$\mathbb{P}[X \geq N] = 1 - \mathbb{P}[X < N] = 1 - \sum_{i=1}^{n-1} \mathbb{P}[X = i]$$

2.2 Partitioning

Non-empty events E_i partition Ω iff:

1. E is *exhausted*, i.e., $\bigcup_i E_i = \Omega$, and
2. E_i (for $i \in [n]$) is *pairwise mutually exclusive*, i.e., $\forall i \neq j.\ E_i \cap E_j = \emptyset$.

2.3 Miscellanea

1. $\mathbb{P}[E] \geq 0$
2. $\mathbb{P}[\Omega] = 1$
3. $E \perp\!\!\!\perp F \to \mathbb{P}[E \cup F] = \mathbb{P}[E] + \mathbb{P}[F]$
4. $\mathbb{P}[E] + \mathbb{P}[E^C] = 1$
5. If $E \subseteq F$, then $\mathbb{P}[E] \leq \mathbb{P}[F]$

2.4 Equally Likely Outcomes

When every outcome in a (finite) sample space Ω is equally likely, and $E \subseteq \Omega$ is an event, then: $\mathbb{P}[E] = |E|/|\Omega|$.

2.5 Conditional Probability

Probability of an event can be *conditioned* on another event, i.e., we compute the probability of E *given* F. This is denoted:

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

We say that events E and F are *independent* iff

$$\mathbb{P}[E \cap F] = \mathbb{P}[E]\mathbb{P}[F] \quad \text{or,}$$
$$\mathbb{P}[E] = \mathbb{P}[E \mid F]$$

2.6 Law of Total Probability

Suppose that $A_1, \ldots, A_n$ is a partition of Ω, and $E \subseteq \Omega$. Then:

$$\mathbb{P}[E] = \sum_{i=1}^{n} \mathbb{P}[E \mid A_i] \cdot \mathbb{P}[A_i]$$

2.7 Bayes' Theorem (& Total Probability)

$$\mathbb{P}[E \mid F] = \frac{\mathbb{P}[F \mid E] \cdot \mathbb{P}[E]}{\mathbb{P}[F]} = \frac{\mathbb{P}[F \mid E] \cdot \mathbb{P}[E]}{\mathbb{P}[F \mid E] \cdot \mathbb{P}[E] + \mathbb{P}[F \mid E^C] \cdot \mathbb{P}[E^C]}$$

2.8 Chain Rule

Suppose that $A_1, \ldots, A_n$ are events. Then:

$$\mathbb{P}\left[\bigcup_{i=1}^{n} A_i\right] = \prod_{i=1}^{n} \mathbb{P}\left[A_i \,\middle|\, \bigcup_{j=1}^{(i-1)} A_j\right]$$

3 Random Variables

A *random variable (r.v.)* is a function $X : \Omega \to \mathbb{R}$ that maps events to numeric values. We say that the *range* of X is denoted Ω_X.

3.1 Probability Functions

If X is discrete, then its probability mass function (where $p_X : \Omega_X \to [0,1]$) is given:

$$p_X(k) = \mathbb{P}[X = k] \quad \text{and} \quad \sum_x p_X(x) = 1$$

Further, if X is discrete, then its *cumulative distribution function* is defined as $F : \mathbb{R} \to [0,1]$, and is given as:

$$F_X(k) = \mathbb{P}[X \leq k] = \sum_{k' \in \Omega_X}^{k} \mathbb{P}[X = k']$$

3.2 Expectation

The *expectation* of a random variable is denoted as $\mathbb{E}[X] = \mu_X$, and is given as:

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

There are several important axioms of expectation:

1. $\mathbb{E}[c] = c,\ c \in \mathbb{R}$
2. $\mathbb{E}[cX] = c \cdot \mathbb{E}[X],\ c \in \mathbb{R}$
3. $\mathbb{E}[aX + bY] = a \cdot \mathbb{E}[X] + b \cdot \mathbb{E}[Y],\ a, b \in \mathbb{R}$

Note also that, if X and Y are independent, that:

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

Note further that, if $Y = f(X)$, then:

$$\mathbb{E}[Y] = \sum_{x \in \Omega_X} g(x) p_X(x)$$

3.3 Variance

The *variance* of a random variable is defined as (where $\mu = \mathbb{E}[X]$):

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$
$$\sigma_X^2 + \mu^2 = \mathbb{E}[X^2]$$

To compute the variance, we use one of (3) tactics:

1. If $|\Omega|$ is small, and/or $\mathbb{E}[X]$ is easy to compute/known, then for each $\omega \in \Omega$, compute $(X - \mathbb{E}[X])^2$ at each point, and take the weighted sum (by ω).
2. Numerically manipulate the second equality above.
3. Use one of the common distributions below.

We say that the standard deviation is given as:

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

Note the following:

$$\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X)$$

When random variables X and Y are independent, (see §2.5), then (and only then) the following holds:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

4 Common Discrete Distributions

4.1 Uniform

We say $X \sim \mathrm{Unif}(a,b)$ when (for $a \leq b$) X has an equally likely chance of taking a value in $[a,b]$. (Example: the outcome of a die roll).

$$p_X(k) = \frac{1}{b-a+1},\ k \in [a,b]$$

$$F_X(x) = \frac{\lfloor x \rfloor - a + 1}{b - a + 1}$$

$$\mathbb{E}[X] = \frac{a+b}{2} \quad \mathrm{Var}(X) = \frac{(b-a)(b-a+2)}{12}$$

4.2 Bernoulli (indicated)

We say $X \sim \mathrm{Ber}(p)$ when X is either 0 or 1 with probability p.

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

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \geq x < 1 \\ 1 & x \geq 1 \end{cases}$$

$$\mathbb{E}[X] = p \quad \mathrm{Var}(X) = p(1-p)$$

4.3 Binomial

We say $X \sim \mathrm{Bin}(n,p)$ when X is the sum of Bernoulli trials. An example of this is the *number* of heads in a series of n coin flips.

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k},\ k \in [0,n]$$

$$F_X(x) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{k} p^i (1-p)^{n-i}$$

$$\mathbb{E}[X] = np \quad \mathrm{Var}(X) = np(1-p)$$

Note that $\mathrm{Bin}(1,p) \triangleq \mathrm{Ber}(p)$. Note further that as $n \to \infty$ and $p \to 0$, that:

$$\lim_{\substack{n \to \infty \\ p \to 0}} \mathrm{Bin}(n,p) = \mathrm{Poi}(\lambda),\ np = \lambda$$

4.4 Geometric

We say that $X \sim \mathrm{Geo}(p)$ when X is the number of Bernoulli trials up to the first successful trials.

$$p_X(k) = (1-p)^{k-1} p,\ k \in \mathbb{N}$$

$$\mathbb{E}[X] = \frac{1}{p} \quad \mathrm{Var}(X) = \frac{1-p}{p^2} \quad F_X(x) = 1 - (1-p)^k$$

4.5 Poisson

We say that $X \sim \mathrm{Poi}(\lambda)$ when X is the number of events during a given time, where λ is the rate of said event.

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!},\ k \in [0, \infty)$$

$$\mathbb{E}[X] = \mathrm{Var}(X) = \lambda$$

Note that if $X_i \sim \mathrm{Poi}(\lambda_i)$, then $X = X_1 + \ldots + X_n \sim \mathrm{Poi}(\lambda_1 + \ldots + \lambda_n)$.

CSE 312 Final – Taylor Blau

5 Conditional Expectation

The Law of Total Expectation states that, if $A_1, \ldots, A_n$ is a total partition, then for a random variable X:

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X \mid A_i] \cdot \mathbb{P}[A_i]$$

Similarly, for random variables X and Y, then:

$$\mathbb{E}[X] = \sum_{y} \mathbb{E}[X \mid Y = y] \cdot p_Y(y)$$

Conditional expectation is defined as follows:

$$\begin{aligned}\mathbb{E}[X \mid Y = y] &= \sum_x x \cdot p_{X|Y}(x \mid y) \\ &= \sum_x x \cdot \mathbb{P}[X = x \mid Y = y]\end{aligned}$$

Likewise, the Linearity of Expectation still holds under conditional expectation:

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

6 Joint Distributions

Define the joint probability mass function as follows:

$$p_{X,Y}(x,y) = \mathbb{P}[X = x \cap Y = y]$$

Where the joint range is denoted $\Omega_{X,Y}$ and is given as:

$$\Omega_{X,Y} = \left\{(x,y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x,y) > 0\right\}$$

We can define the *marginal probability mass function* as follows:

$$p_X(x) = \sum_{y \in \Omega_Y} p_{X,Y}(x,y)$$

Or by "summing the rows".

We define *joint expectation of a function* as follows:

$$\mathbb{E}[g(X,Y)] = \sum_{(x,y)\in\Omega_{X,Y}} g(x,y) \cdot p_{X,Y}(x,y)$$

Further, we define the *joint cumulative density function* for discrete random variables as:

$$\begin{aligned}F_{X,Y}(x,y) &= \mathbb{P}(X \le x, Y \le y) \\ &= \sum_{s \le x, t \le y} p_{X,Y}(s,t)\end{aligned}$$

Lastly, we define *independence of joint discrete random variables* as having been met iff:

$$\forall (x,y) \in \Omega_X \times \Omega_Y.\, p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$$

7 Continuous Random Variables

A *cumulative distribution function* (cdf) is given for a random variable X as $F_X(x) = \mathbb{P}(X \le x)$, where F is a monotone non-decreasing function from $0 \to 1$.

A *continuous random variable* X is given when $F_X : \mathbb{R} \to \mathbb{R}$, and is continuous everywhere. Ω has an un-countably infinite quantity of values.

7.1 Univariate Continuous R.V.'s

A *probability density function* is given for a random variable X as $f_X : \mathbb{R} \to \mathbb{R}$, and is defined as:

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

From the above, it follows that:

$$\mathbb{P}[a \le X \le b] = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$$

Note also that:

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Critically, note that $f_X(a) \neq \mathbb{P}(X = a)$, since $\mathbb{P}(X = a) = 0$. This follows from the fact that the universe Ω is uncountable infinite, thus the probability of any *individual* event occurring tends towards 0.

Lastly, note that the expectation of a function g over a continuous random variable X is given as:

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

For an event A and continuous random variable X the *Law of Total Probability* is given as:

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

Likewise, for continuous random variable Y, the *Law of Total Expectation* is given as follows:

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

7.2 Operations on R.V.'s

Let X be a random variable, with $\mathbb{E}[X] = \mu$ and $\text{Var}(X) = \sigma^2$. It follows that:

$$Y = \frac{X - \mu}{\sigma} \Rightarrow \mathbb{E}[Y] = 0 \text{ and } \text{Var}(Y) = 1$$

The normal distribution is "closed", or that if $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

Normal variables are also "reproductive", i.e., that if $X_1, \ldots, X_n$ are independent normal random variables with $\mathbb{E}[X_i] = \mu_i$ and $\text{Var}(X_i) = \sigma_i^2$, and $a_1, \ldots, a_n \in \mathbb{R}$, and $b \in \mathbb{R}$, then:

$$\begin{aligned}X &= \sum_{i=1}^{n}(a_i X_i + b) \\ &\sim \mathcal{N}\left(\sum_{i=1}^{n}(a_i\mu_i + b), \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)\end{aligned}$$

7.3 Multivariate Continuous R.V.'s

Denote the *joint probability density function* of a continuous random variable as $f_{X,Y}(x,y)$. Note that (for reasons above):

$$f_{X,Y}(x,y) \neq \mathbb{P}(X = x \cap Y = y)$$

Similarly, denote the *joint cumulative density function* of a continuous random variable as $F_{X,Y}(x,y)$, where:

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\, ds\, dt$$

The "normalization" property still holds, too:

$$\iint_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy = 1$$

We define the *marginal probability density function* of two continuous random variables, X and Y as:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

Analogous to the discrete setting, we define the *joint expectation of a function* for continuous random variables X, Y and function $g : \mathbb{R}^2 \to \mathbb{R}^2$ as:

$$\mathbb{E}[g(X,Y)] = \iint_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dx\, dy$$

Carrying on to the *conditional probability density function* for random variables X and Y, it is given as:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

The *joint expectation* of continuous random variables X, and Y is defined as:

$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x,y)\, dx$$

8 Common Continuous Distributions

8.1 Uniform

We say that if $X \sim \text{Unif}_C(a,b)$ then X has an equally likely chance of taking any real value between a and b.

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

$$\mathbb{E}[X] = \frac{a+b}{2} \quad \text{and} \quad \text{Var}(X) = \frac{(b-a)^2}{12}$$

8.2 Exponential

We say that if $X \sim \text{Exp}(\lambda)$ that:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = 1 - e^{-\lambda x},\ x \ge 0$$

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \text{and} \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

Note that this is the continuous analog of the geometric discrete random variable, representing the *waiting time* until the next event, where λ is the avg. number of events per unit time. The Poisson measures a quantity of events, unlike the Exponential, which measures the waiting time.

The Exponential random variable is "memoryless", i.e., that for any $s, t \ge 0$:

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

8.3 Normal

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then:

$$F_X(x) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\, dx = \Phi(x)$$

$$\mathbb{E}[X] = \mu \quad \text{and} \quad \text{Var}(X) = \sigma^2$$

The "standard normal" random variable is usually denoted $Z \sim \mathcal{N}(0,1)$. Note also that:

$$\Phi(-x) = 1 - \Phi(x)$$

9 CLT, Tail Bounds, & Likelihood

9.1 Central Limit Theorem

If $X_1, \ldots, X_n$ are i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. Thus, $\mathbb{E}[X] = n\mu$ and $\text{Var}(X) = n\sigma^2$. Define $\bar{X}$ to be $\frac{1}{n}\sum_{i=1}^{n} X_i$, where $\mathbb{E}\left[\bar{X}\right] = \mu$ and $\text{Var}\left(\bar{X}\right) = \frac{1}{n}\sigma^2$. Then:

$$\lim_{n\to\infty} \bar{X} \sim \mathcal{N}\left(\mu, \frac{1}{n}\sigma^2\right)$$

9.2 Tail Bounds: Markov's Inequality

Let X be a non-negative random variable, and $\alpha > 0$. Then, Markov's Inequality states that:

$$\mathbb{P}(X \ge \alpha) \le \frac{1}{\alpha}\mathbb{E}[X]$$

9.3 Tail Bounds: Chebyshev's Inequality

Let Y be a non-negative random variable with $\mathbb{E}[Y] = \mu$ and $\text{Var}(Y) = \sigma^2$. Then, for $\alpha > 0$:

$$\mathbb{P}(|Y - \mu| \ge \alpha) \le \frac{\sigma^2}{\alpha^2}$$

9.4 Law of Large Numbers

Weak Law $\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| > \epsilon) = 0$

Strong Law $\mathbb{P}\left(\lim_{n\to\infty} \bar{X}_n \mu\right) = 1$

9.5 Likelihood Functions, Estimators

The Likelihood Function is given as:

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

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

And combines as the product of likelihood functions; that is to say the And combines as the product of likelihood functions; that is to say the Likelihood is a measure of how well a sampling of data, say $\vec{x}$, *fits* to an assumed distribution given parameters $\vec{\theta}$.

To compute the *maximum likelihood estimator*, we compute the maximum as follows. Say we are interested in θ_i, then:

$$0 = \frac{\partial}{\partial\theta_1} \log L \bigg|_{\theta = \widehat{\theta_1}}$$

9.6 Bias of Likelihood Estimators

The *bias* of an estimator is how far it deviates from actuality:

$$\text{Bias}\left(\widehat{\theta}, \theta\right) = \mathbb{E}[\widehat{\theta}] - \theta$$