Review of Math 461
An Introduction to Mathematical Methods of Probability

Alex Thies

Abstract. The following is a brief review of the topics covered in Math 461 - An Introduction to Mathematical Methods of Probability, taught by Professor David Levin at the University of Oregon during the Fall Quarter of 2016.

Contents

1. Axioms of Probability

(1) ***Axiom 1*** Let $E \subsetneq S$ be an event within the sample space S,
$$0 \leq P(E) \leq 1$$
(2) ***Axiom 2*** Let S be a sample space,
$$P(S) = 1$$
(3) ***Axiom 3*** For any sequence of mutually exclusive events $E_1, E_2, \dots$ (that is, events for which $E_i E_j = \emptyset$ when $i \neq j$),
$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P\left(E_i\right)$$
We refer to $P(E)$ as the probability of the event E.

2. Combinatorics

2.1. Counting.

2.2. Permutations.

$$n! = n(n-1)(n-2)\cdots(n-n+1)$$

2.2.1. *Stirling's Approximation of $n!$.*
$$n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n}$$

2.3. Binomial Coefficients. Binomial coefficients describe the number of possible ways that we can choose k objects from a group of n objects. The classic example is the ways to pull colored objects from a container.
$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$$

2.4. Multinomial Coefficients.

$$\binom{n}{a_1\ a_2\ a_3\ \dots\ a_k} = \frac{n!}{a_1! \cdot a_2! \cdot a_3!\ \cdots\ a_k!}$$

2.5. Conditional Probability.

$$P(A|B) = \frac{P(AB)}{P(B)},$$
$$P(AB) = P(A|B)P(B)$$

3. Conditional Probability & Independence

3.1. Bayes' Formula.

$$\begin{aligned} P(A) &= P(AB) + P(AB^c), \\ P(AB) &= P(AB)P(B) + P(AB^c)P(B^c), \\ &= P(AB)P(B) + P(AB^c)\left[1 - P(B)\right]. \end{aligned}$$

3.2. Inclusion-Exclusion Principle.

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) - \sum_{1 \leq i < j \leq \infty} P(A_i A_j) + \sum_{1 \leq i < j < k \leq \infty} P(A_i A_j A_k) \pm \dots$$

3.3. Independence.

4. Discrete Random Variables

4.1. General Principles. Let X be a discrete random variable, then

(1) The Probability Mass Function (pmf) is a function which describes the probability that a random variable X outputs a result x, i.e., $p_X(x) = P(X = x)$. The pmf for a random variable can be written in discrete terms, or formulaically,

(a) Discrete number of terms,

$$\begin{aligned} p_X(x_1) &= p_1, \\ p_X(x_2) &= p_2, \\ &\vdots \\ p_X(x_i) &= p_i, \\ &\vdots \\ p_X(x_n) &= p_n. \\ p_X(x) &= 0, \text{ for all other } x. \end{aligned}$$

The above equations, taken together, are a pmf for X. Note that $\sum_{i=1}^{n} P_X(x_i) = 1$, this comes from the Axioms of Probability.

(b) We can also express the pmf as a formula, generally speaking this is, $p_X(x) = f(x)$. The easiest (and most important) example is the pmf for a binomial distribution, that is, for $X \sim \text{Binomial}(p)$,

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Note that $\sum_{i=1}^{n} p_X(x_i) = 1$.v

(2) The following function outputs the Expectation, or mean μ of a discrete random variable X,

$$E(X) = \sum_{x \in X} x p_X(x)$$

If we want to compute the expecation of a function f of X, we perform the following computation,

$$E(f(X)) = \sum_{f(x) \in f(X)} f(x) p_X(x)$$

Note that the argument of the pmf is unchanged. For example,

$$\begin{aligned} E\left(X^2\right) &= \sum_{x^2 \in X^2} x^2 \cdot p_X(x) \\ E(1/X) &= \sum_{x^{-1} \in X^{-1}} \frac{1}{x} \cdot p_X(x) \end{aligned}$$

(3) The Variance of a random variable is the expectation of the squared difference between the expectation and the result. The following function

outputs the variance, σ^2, of a discrete random variable X,

$$\begin{aligned}\operatorname{Var}(X) &= E\left[(X-\mu)^2\right], \text{ or, more simply,}\\ \operatorname{Var}(X) &= E(X^2) - [E(X)]^2.\end{aligned}$$

The following are properties of the Variance,

$$\operatorname{Var}(aX+b) = a^2 \operatorname{Var}(X).$$

(4) The Standard Deviation of a random variable is the square root of the variance, i.e., $\operatorname{SD}(X) = \sqrt{\operatorname{Var}(X)} = \sqrt{\sigma^2} = \sigma$.

4.2. Specific Discrete Random Variables.

4.2.1. *Binomial.* A binomial random variable X has binary outputs, e.g., success or failure, with p denoting the probability of success, and $(1-p)$ denoting the probability of failure, in n trials; n trials and probability of success p are the parameters for X, hence, we write $X \sim \text{Binomial}(n,p)$. The standard example for a binomial random variable is tossing a coin n times, with the probability of the coin landing on heads being p.

(1) Probability Mass Function. The pmf of $X \sim \text{Binomial}(n,p)$ is

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

(2) Expectation. The expectation of a binomial random variable is the product of the number of trials and the probability of success, i.e., for $X \sim \text{Binomial}(n,p)$, $E(X) = np$. Before we prove this, note that

$$x\binom{n}{x} = n\binom{n-1}{x-1} = n\left(\frac{(n-1)!}{(n-x-1)!(x-1)!}\right).$$

We compute the following,

$$
\begin{aligned}
E(X^k) &= \sum_{x^k \in X^k} x^k p_X(x), \\
&= \sum_{x=1}^{n} x^k \binom{n}{x} p^x (1-p)^{n-x}, \\
&= \sum_{x=1}^{n} n x^{k-1} \binom{n-1}{x-1} p^x (1-p)^{n-x}, \\
&= np \sum_{x=1}^{n} x^{k-1} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x}, \\
&= np \sum_{x=1}^{n} x^{k-1} \left(\frac{(n-1)!}{(n-(x-1))!(x-1)!} \right) p^{x-1} (1-p)^{n-x}, \\
&\text{if we set } y = x-1, \\
&= np \sum_{y=0}^{n-1} (y+1)^{k-1} \left(\frac{(n-1)!}{(n-y)!y!} \right) p^{y} (1-p)^{(n-1)-y}, \\
&= np \sum_{y=0}^{n-1} (y+1)^{k-1} \binom{n-1}{y} p^{y} (1-p)^{(n-1)-y}, \\
&= np \cdot E\left[(Y+1)^{k-1} \right]
\end{aligned}
$$

If we let $k = 1$, then we see that for $X \sim \text{Binomial}(n, p)$, $E(X) = np$.

(3) Variance. The variance for a binomial random variable is the product of the number of trials, the probability of success, and the probability of failure. Recall that $\text{Var}(X) = E(X^2) - [E(X)]^2$, thus we must compute $E(X^2)$. From our previous computations we know that if we set $k = 2$, then,

$$
\begin{aligned}
E(X^2) &= np \cdot E\left(Y+1\right), \\
&= np\left[(n-1)p + 1\right], \\
&= np(np - p + 1), \\
&= n^2p^2 - np^2 + np.
\end{aligned}
$$

We now compute the variance,

$$
\begin{aligned}
\text{Var}(X) &= E(X^2) - \left[E(X)\right]^2, \\
&= \left(n^2p^2 - np^2 + np\right) - (np)^2, \\
&= np - np^2, \\
&= np(1-p).
\end{aligned}
$$

(4) Standard Deviation. The standard deviation for a binomial random variable is as follows

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

4.2.2. *Bernoulli.*

(1) Probability Mass Function. A bernoulli random variable X has binary outputs for 1 trial, e.g., success or failure, with p denoting the probability of success, and $(1-p)$ denoting the probability of failure. Bernoulli random variables are the special case of a binomial random variable where $n = 1$, e.g., $\text{Binomial}(1, p) \sim X \sim \text{Bernoulli}(p)$
(2) Expectation. The expectation of a bernoulli random variable is its probability p, i.e., $E(X) = p$. Note that $X \sim \text{Bernoulli}(p) \Rightarrow X \sim \text{Binomial}(1, p)$, thus

$$\begin{aligned} E(X) &= np, \\ &= (1)p, \\ &= p. \end{aligned}$$

(3) Variance. The variance of a bernoulli random variable is its probability p, i.e., $\text{Var}(X) = p$. This is verified from the variance of a binomial random variable with $n = 1$, as we did for the expectation.
(4) Standard Deviation. The standard deviation of a bernoulli random variable is the square root of its probability p, i.e., $\text{SD}(X) = \sqrt{p}$.

4.2.3. *Poisson.* A poisson random variable BLAH. Importantly, a poisson random variable approximates a binomial random variable for large n and small p; the product of these values, $\lambda = np$, is the parameter for a poisson random variable, i.e., $X \sim \text{Poisson}(\lambda)$.

(1) Probability Mass Function. The pmf of $X \sim \text{Poisson}(\lambda)$ is

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

(2) Expectation. The expectation of a poisson random variable is its parameter $\lambda = np$, for n trials and probability of success p. We compute the following,

$$\begin{aligned} E(X) &= \sum_{x \in X} x p_X(x), \\ &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!}, \\ &= \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}, \\ &= \lambda \sum_{y=1}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!}. \end{aligned}$$

Note that $\frac{e^{-\lambda}\lambda^y}{y!}$ is the probability mass function for $Y \sim \text{Poisson}(\lambda)$, thus, from the axioms of probability, the summation of the pmf over its domain is equal to 1, therefore $E(X) = \lambda$.

(3) Variance. The variance of a poisson random variable is its parameter $\lambda = np$. We compute the following,

$$\begin{aligned}
\mathrm{Var}(X) &= E\left(X^2\right) - \left[E(X)\right]^2, \\
&= \left(\sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!}\right) - \lambda^2, \\
&= \left(\lambda \sum_{x=1}^{\infty} \frac{xe^{-\lambda}\lambda^{x-1}}{(x-1)!}\right) - \lambda^2, \\
&= \lambda \left(\sum_{y=0}^{\infty} \frac{(y+1)e^{-\lambda}\lambda^y}{y!}\right) - \lambda^2, \\
&= \lambda \left(\sum_{y=0}^{\infty} \frac{ye^{-\lambda}\lambda^y}{y!} + \sum_{y=1}^{\infty} \frac{e^{-\lambda}\lambda^y}{y!}\right) - \lambda^2, \\
&= \lambda\left(\lambda+1\right) - \lambda^2, \\
&= \lambda.
\end{aligned}$$

(4) Standard Deviation. The standard deviation for a poisson random variable is as follows

$$\mathrm{SD}(X) = \sqrt{\lambda}.$$

5. Continuous Random Variables

5.1. General Principles.

(1) Probability Density Function.
(2) Cumulative Distribution Function. The cumulative distribution function (cdf) is a function which outputs the probablity that the result of a continuous random variable X is less than a fixed value of $x \in X$, i.e.,

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\ dt,$$

where f_X is a pdf.
(3) Relationship between the pdf and the cdf.
(4) Expectation. The following function outputs the Expectation, or mean μ of a continous random variable X,

$$E(X) = \int_{-\infty}^{\infty} x f_X(x).$$

Note that this case is perfectly analagous to that of the discrete case, as an integral over the real line is equivalent to the summation over the domain of X.

If we want to compute the expecation of a function g of X, we perform the following computation,

$$E\left(f(X)\right) = \int_{-\infty}^{\infty} g(x) f_X(x)$$

Note that the argument of the pmf is unchanged. For example,

$$E\left(X^2\right) = \sum_{x^2 \in X^2} x^2 \cdot p_X(x)$$
$$E\left(1/X\right) = \sum_{x^{-1} \in X^{-1}} \frac{1}{x} \cdot p_X(x)$$

(5) Variance. The Variance of a continuous random variable is the same as that for the discrete case. The following function outputs the variance, σ^2, of a continuous random variable X,

$$\operatorname{Var}(X) = E\left[(X-\mu)^2\right], \text{ or, more simply,}$$
$$\operatorname{Var}(X) = E(X^2) - \left[E(X)\right]^2.$$

(6) Standard Deviation. The Standard Deviation of a random variable is the square root of the variance, i.e., $\mathrm{SD}(X) = \sqrt{\operatorname{Var}(X)} = \sqrt{\sigma^2} = \sigma$.

5.2. Specific Continuous Random Variables.

5.2.1. *Normal.*

(1) Probability Density Function. The pdf for a normal random variable is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

(2) Cumulative Distribution Function. The cdf for a normal random variable is $\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-(t-\mu)^2/2\sigma^2}\, dt$, note that this generalizes to the case $\int_{-\infty}^{x} e^{-t^2}\, dt$, for which there is no anti-derivative.

(3) Expectation.

$$E(X) = \mu$$

(4) Variance.

$$\operatorname{Var}(X) = \sigma^2$$

(5) Standard Deviation.

$$\mathrm{SD}(X) = \sigma$$

5.2.2. *Standard Normal.* The standard normal random variable is a normal random variable centered about $x = 0$ with standard deviation 1, i.e., $X \sim \mathrm{Normal}(0,1)$, $\mu = 0$, and $\sigma^2 = 1$.

(1) Probability Density Function. The pdf for a normal random variable is

$$f_X(x) = \frac{e^{-x^2}}{\sqrt{2\pi}}$$

(2) Cumulative Distribution Function. For a normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 1$, the cdf is tabled, and denoted by the $\Phi(x)$ function. Thus, $F_X(2) = P(X < 2) = \Phi(2)$.

5.2.3. *Exponential.*

(1) Probability Density Function.

$$f_X(x) = \begin{cases} \lambda e^{\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(2) Cumulative Distribution Function.

$$\begin{aligned} F_X(x) &= P(X < x), \\ &= \int_{-\infty}^{x} \lambda e^{-\lambda t}\, dt, \\ &= -e^{-\lambda t}\Big|_0^x, \\ &= 1 - e^{-\lambda x} \quad x \geq 0 \end{aligned}$$

(3) Expectation.

$$\begin{aligned} E(X^n) &= \int_0^\infty x^n \lambda e^{-\lambda x}\, dx, \\ &= -x^n e^{-\lambda x}\big|_0^\infty + \int_0^\infty e^{-\lambda x} n x^{n-1}\, dx, \\ &= 0 + \frac{n}{\lambda}\int_0^\infty \lambda e^{-\lambda x} x^{n-1}\, dx, \\ &= \frac{n}{\lambda} E(X^{n-1}) \end{aligned}$$

Letting $n = 1$ shows that $E(X) = 1/\lambda$.

(4) Variance. From the above computation of $E(X^n)$, if we let $n = 2$ we see that $E\left(X^2\right) = \frac{2}{\lambda}E(X) = \frac{2}{\lambda^2}$, thus the variance is,

$$\begin{aligned} \mathrm{Var}(X) &= E\left(X^2\right) - \left[E\left(X\right)\right]^2, \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2}, \\ &= \frac{1}{\lambda^2} \end{aligned}$$

(5) Standard Deviation. The Standard Deviation of a continuous random variable is,

$$\mathrm{SD}(X) = \frac{1}{\lambda}$$

5.2.4. *Uniform.*

(1) Probability Density Function.

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

(2) Cumulative Distribution Function.

$$\begin{aligned} F_X(x) &= P(X < x) \\ &= \begin{cases} 1 & x > b \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 0 & x < a \end{cases} \end{aligned}$$

(3) Expectation.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x)\ dx, \\ &= \int_a^b \frac{x}{b-a}\ dx, \\ &= \frac{b^2-a^2}{2(b-a)}, \\ &= \frac{b+a}{2} \end{aligned}$$

(4) Variance.

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x)\ dx, \\ &= \int_a^b \frac{x^2}{b-a}\ dx, \\ &= \frac{b^3-a^3}{3(b-a)}, \\ &= \frac{b^2+ab+a^2}{3} \end{aligned}$$

Thus,

$$\begin{aligned} \mathrm{Var}(X) &= E\left(X^2\right) - \left[E\left(X\right)\right]^2, \\ &= \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2, \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

(5) Standard Deviation.

$$\mathrm{SD}(X) = \frac{b-a}{2\sqrt{3}}$$

6. Jointly Distributed Random Variables

6.1. General Principles.

(1) Joint Mass Function
(2) Marginal Mass Functions
(3) Joint Density Function

$$f_{X,Y}(x,y) =$$

(4) Marginal Density Functions

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy, \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx. \end{aligned}$$

(5) Cumulative Distribution Function

$$\begin{aligned} F_{X,Y}(x,y) &= P(X < x,\ Y < y), \\ &= \int_{t=-\infty}^{y} \int_{s=-\infty}^{x} f_{X,Y}(s,t)\ dsdt \end{aligned}$$

(6) Expectation

$$\begin{aligned} E(XY) &= E(X)E(Y), \\ &= \int_{-\infty}^{\infty} x f_X(x)\ dx \cdot \int_{-\infty}^{\infty} y f_Y(y)\ dy, \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dydx \cdot \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dxdy, \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x,y)\ dxdy \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x,y)\ dxdy, \\ &= \iint_{-\infty}^{\infty} xy f_{X,Y}(x,y)\ dxdy. \end{aligned}$$

$E(X+Y) = E(X) + E(Y)$

(7) Covariance

$$\operatorname{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2).$$

$$\begin{aligned} \operatorname{Cov}(X,X) &= \operatorname{Var}(X), \\ \operatorname{Cov}(X,Y) &= \operatorname{Cov}(Y,X), \\ \operatorname{Cov}(X+Y,Z) &= \operatorname{Cov}(X,Z) + \operatorname{Cov}(Y,Z), \\ \operatorname{Cov}(aX+b,Y) &= a\operatorname{Cov}(X,Y) \end{aligned}$$

(8) Correlation

$$\operatorname{Corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\operatorname{SD}(X)\operatorname{SD}(Y)}$$

7. Theorems

7.1. The Law of Large Numbers.

7.1.1. *The Weak Law of Large Numbers.* Let $X_1, X_2, \dots$ be sequence of independent and identically distributed random variables, each having finite mean $E(X_i) = \mu$. Then, for any $\epsilon > 0$,

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0 \quad as \quad n \to \infty$$

7.1.2. *The Strong Law of Large Numbers.* Let $X_1, X_2, \dots$ be sequence of independent and identically distributed random variables, each having finite mean $\mu = E(X_i)$. Then, with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \quad as \quad n \to \infty$$

7.2. Central Limit Theorem.

7.2.1. *The DeMoivre-Laplace Limit Theorem.* If S_n denotes the number of successes that occur when n independent trials, each resulting in a success with probability p, are performed, then, for any $a < b$,

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \to \Phi(b) - \Phi(a)$$

as $n \to \infty$.

7.2.2. *The Central Limit Theorem.* Let $X_1, X_2, \dots$ be sequence of independent and identically distributed random variables, each having mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$,

$$P\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx \quad as \quad n \to \infty$$

8. Applications

8.1. Probability density functions for functions of continuous random variables. Given a pdf $f_X(x)$, find the pdf of a function of X, i.e., for $Y = g(X)$, find $f_Y(y)$

8.2. Tranformations of jointly distributed random variables.

8.3. Z-Scores.

9. Skipped Materials

9.1. Markov's Inequality. If X is a random variable that takes only non-negative values, then, for any value $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}$$

9.2. Chebyshev's Inequality. If X is a random variable with finite mean μ and variance σ^2, then, for any value $k > 0$,

$$P\left(|X - \mu| \geq k\right) \leq \frac{\sigma^2}{k^2}$$

9.3. Markov Chains.

10. Classic Problems

10.1. Coupons.

Department of Mathematics, University of Oregon, Eugene, OR
E-mail address: athies@uoregon.edu