1 Review

We begin our discussion of mathematical statistics with a review of concepts from previous courses. One of these key concepts is that of probability.

Recall that the **sample space** Ω is the set of all possible outcomes of an experiment. Subsets of Ω are called **events** and the collection of all events is denoted by $\mathcal{F}$.

Definition 1.1 (probability set function)**.** Let Ω be a sample space and let $\mathcal{F}$ be the collection of all events. Let P be a real-valued function defined on $\mathcal{F}$. Then P is a **probability set function** (also referred to as **probability measure**, **probability distribution** or simply **probability**) if it satisfies the following three conditions:

1. $0 \leq P(A) \leq 1$, for all $A \in \mathcal{F}$.
2. $P(\Omega) = 1$ and $P(\emptyset)$.
3. If $\{A_n\}$ is a sequence of events in $\mathcal{F}$ and $A_m \cap A_n = \emptyset$ for all $m \neq n$, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Another important concept is that of a random variable and its support in order to formalize quantities which depend on random events.

Definition 1.2 (random variable)**.** Let Ω be a sample space. A **random variable**, often abbreviated RV, is a function from Ω into the real numbers. The **support** (also called **space** or **range**) of X is the set of real numbers $\mathcal{S} = \{x : x = X(\omega), \omega \in \Omega\}$.

In cases where $\mathcal{S}$ is a countable set, we say that X is a **discrete RV**. The set $\mathcal{S}$ may also be an interval of real numbers, in which case we say that X is a **continuous RV**.

Given a random variable X, its support $\mathcal{S}$ becomes the sample space of interest. Besides inducing the sample space $\mathcal{S}$, X also induces a probability which we call the **distribution** of X.

The probability distribution of a discrete random variable is described completely in terms of its probability mass function and its support.

Definition 1.3 (pmf)**.** Let X be a discrete random variable with support $\mathcal{S}$. The **probability mass function** (pmf) of X is given by

$$p_X(x) = P(X = x), \text{ for } x \in \mathcal{S}.$$

Similarly, the probability distribution of a continuous random variable is described completely in terms of its probability density function and its support.

Definition 1.4 (pdf)**.** Let X be a continuous random variable with support $\mathcal{S}$. The **probability density function** (pdf) of X is a function f_X that satisfies

$$P(X \in A) = \int_{\mathcal{S}} f_X(x)\,\mathrm{d}x$$

where A is a subset of $\mathbb{R}$ that can be written as a countable union of intervals.

The pmf of a discrete random variable and the pdf of a continuous random variable are quite different entities. The cumulative distribution function, though, uniquely determines the probability distribution of a random variable.

Definition 1.5 (cdf)**.** Let X be a random variable. Then its **cumulative distribution function** (cdf) is defined by $F_X(x)$, where

$$F_X(x) = P(X \le x).$$

One of the most important measures associated with RVs is that of expectation.

Definition 1.6 (expectation)**.** Let X be a random variable with support $\mathcal{S}$.
If X is a *continuous* random variable with pdf $f(x)$ and

$$\int_{\mathcal{S}} |x| f(x)\,\mathrm{d}x$$

is finite, then the **expectation** of X, denoted $E(X)$ is defined as

$$E(X) = \int_{\mathcal{S}} x f(x)\,\mathrm{d}x\,.$$

If X is a *discrete* random variable with pmf $p(x)$ and

$$\sum_{x\in\mathcal{S}} |x| p(x)$$

is finite, then the **expectation** of X is defined as

$$\sum_{x\in\mathcal{S}} x p(x).$$

Sometimes the expectation $E(X)$ is called the **expected value** of X or the **mean** of X. When the mean designation is used, we often denote the expected value by μ.

Theorem 1.1. Let X be a random variable and let $Y = g(X)$ for some function g.

(a) Suppose X is discrete with pmf $p_X(x)$ and support $\mathcal{S}_X$. If $\sum_{x\in\mathcal{S}_X} |g(x)| p_X(x)$ is finite, then the expectation of Y exists and is given by

$$\sum_{x\in\mathcal{S}_X} g(x) p_X(x).$$

(b) Suppose X is continuous with pdf $f_X(x)$ and support $\mathcal{S}_X$. If $\int_{\mathcal{S}_X} |g(x)| f_X(x)\,\mathrm{d}x$ is finite, then the expectation of Y exists and is given by

$$\int_{\mathcal{S}_X} g(x) f_X(x)\,\mathrm{d}x\,.$$

An important application of the above theorem shows that expectation is *linear*. That is, $E(aX+b) = aE(X)+b$. Furthermore, this property can be generalized for $a_1,\ldots,a_k$ real numbers and

$g_1, \dots, g_k$ real-valued functions.

$$E(a_1 g_1(X) + \cdots + a_k g_k(X)) = a_1 E(g_1(X)) + \cdots + a_k E(g_k(X))$$

Expectation allows us to define a countably infinite number of measures associated with RVs, called moments.

Definition 1.7 (moment)**.** Suppose X is a RV and m is a positive integer. The ***m*th moment** of X is defined to be $E(X^m)$, provided this expectation exists.

As such, the 1st moment of a RV is simply its **mean** μ. It is often useful to think about moments about the mean $E((X-\mu)^m)$. We call these **central moments**. The 2nd central moment should be familiar to you as the **variance** σ^2. We call the 3rd central moment the **skewness** and call the 4th central moment the **kurtosis**.

Definition 1.8 (moment generating function)**.** Let X be a random variable such that for some $h > 0$, the expectation of e^{tX} exists for $-h < t < h$. The **moment generating function** of X is defined to be the function $M_X(t) = E(e^{tX})$ for $-h < t < h$.

Clearly, $M_X(0) = 1$ for any RV. Not every random variable has a mgf. For example, the mgf of the Cauchy Distribution with pdf $f(x) = \frac{1}{\pi(1+x^2)}$ is not defined. It can be shown that if the mgf of a RV exists, then all of its moments exist.

Theorem 1.2. Let X and Y be RVs with mgfs M_X and M_Y, respectively, existing in open intervals about 0. Then $F_X(z) = F_Y(z)$ for all $z \in \mathbb{R}$ if and only if $M_X(t) = M_Y(t)$ for all $t \in (-h, h)$ for some $h > 0$.

Theorem 1.3. Let X be a RV with mgf M_X, and let $a, b \in \mathbb{R}$ be fixed. Then the mgf of $Y = aX + b$ also exists and is given by

$$M_Y(t) = e^{bt} M_X(at).$$

Theorem 1.4. Suppose X and Y are independent RVs with mgfs M_X and M_Y. Let $a, b \in \mathbb{R}$ be fixed and define $Z + aX + bY$. Then the mgf of Z exists in an open interval about 0 and is given by

$$M_Z(t) = M_X(at) M_Y(bt).$$

Theorem 1.5. Suppose X is a RV with mgf M_X and $M_X^{(m)}(t) = \frac{\mathrm{d}^m}{\mathrm{d}t^m} M_X(t)$. Then the mth moment of X is given by

$$E(X^m) = M_X^{(m)}(0).$$

The above theorem should make clear why we call mgfs as such. The proof is reliant on the Taylor expansion of e^{tX}. Observe the following.

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= E\left(\sum_{n=0}^{\infty} \frac{t^n}{n!} X^n \right) \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n) \end{aligned}$$

We reintroduce some special distributions, starting with those of the discrete kind.

Definition 1.9 (binomial RV)**.** Assume a sequence of n Bernouilli trials each with probability of success p and let X be the number of successes. Then X is a **binomial RV** with pmf

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

We write $X \sim b(n, p)$.

If $X \sim b(n, p)$, X has support $\{0, 1, \ldots, n\}$, mean $\mu = np$, variance $\sigma^2 = np(1-p)$ and mgf $M_X(t) = (1 - p + pe^t)^n$.

Definition 1.10 (negative binomial RV)**.** Assume a sequence of Bernouilli trials each with probability of success p is performed until the rth success occurs. Let Y be the number of trials required. Then Y is a **negative binomial RV** with pmf

$$p_Y(y) = \begin{cases} \binom{y+r-1}{y-1} p^r (1-p)^y & \text{for } y = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

We write $Y \sim nb(r, p)$.

If $Y \sim nb(r, p)$, Y has support $\mathbb{Z}_{\geq 0}$, mean $\mu = \frac{pr}{1-p}$, variance $\sigma^2 = \frac{pr}{(1-p)^2}$ and mgf $M_Y(t) = \left(\frac{1-p}{1-pe^t} \right)$ with $t < -\ln p$.

Taking $r = 1$, we obtain the geometric distribution.

Definition 1.11 (Poisson RV)**.** A discrete RV X is a **Poisson RV** if its pmf has the form

$$p_X(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{for } x = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $\lambda \in \mathbb{R}_{\geq 0}$. We write $X \sim \text{Pois}(\lambda)$.

If $X \sim \text{Pois}(\lambda)$, X has support $\mathbb{Z}_{\geq 0}$, mean $\mu = \lambda$, variance $\sigma^2 = \lambda$ and mgf $M_X(t) = \exp(\lambda(e^t - 1))$.

We now recall some continuous distributions.

Definition 1.12 (uniform RV)**.** A continuous RV X is said to be a **uniform RV** if it has pdf

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

where $a, b \in \mathbb{R}$ are fixed. We write $X \sim U(a, b)$.

If $X \sim U(a, b)$, X has support $[a, b]$, mean $\mu = \frac{a+b}{2}$, variance $\sigma^2 = \frac{(b-a)^2}{12}$ and mgf $M_X(t) = \begin{cases} \frac{e^{tb} - e^{ta}}{t(b-a)} & \text{for } t \neq 0 \\ 1 & \text{for } t = 0 \end{cases}$.

Definition 1.13 (normal RV)**.** A continuous RV X is said to be a **normal RV** with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ if its pdf has the form

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

We write $X \sim N(\mu, \sigma^2)$.

If $X \sim N(\mu, \sigma^2)$, X has support $\mathbb{R}$, mean μ, variance σ^2 and mgf $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$.

The derivation of this mgf follows, as a review.

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{tx} \, \mathrm{d}x
\end{aligned}
$$

Let $z = \frac{x-\mu}{\sigma}$.

$$
\begin{aligned}
&= e^{\mu t} \int e^{z\sigma t} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}z^2} \, \mathrm{d}z \\
&= e^{\mu t} e^{\frac{\sigma^2 t^2}{2}} \\
&= e^{\mu t + \frac{\sigma^2 t^2}{2}}
\end{aligned}
$$

Theorem 1.6. Let $X_1, \ldots, X_n$ be IID RVs such that $X_i \sim N(\mu_i, \sigma_i^2)$ for each $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^n a_i X_i$ for some set of real constants $\{a_1, \ldots, a_n\}$. Then Y is also normally distributed and

$$
Y \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right).
$$

Definition 1.14 (gamma RV)**.** A continuous RV X is said to be a **gamma RV** with parameters $\alpha, \beta > 0$ if its pdf has the form

$$
f_X(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{for } x \in \mathbb{R}_{>0} \\ 0 & \text{otherwise} \end{cases}.
$$

We write $X \sim \Gamma(\alpha, \beta)$.

Taking $\alpha = 1$ yields the exponential distribution.

If $X \sim \Gamma(\alpha, \beta)$, X has support $\mathbb{R}_{>0}$, mean $\mu = \frac{\alpha}{\beta}$, variance $\sigma^2 = \frac{\alpha}{\beta^2}$ and mgf $M_X(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha}$ for $t < \beta$.

Definition 1.15 (beta RV)**.** A continuous RV X is said to be a **beta RV** with parameters $\alpha, \beta > 0$ if its pdf has the form

$$f_X(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} & \text{for } x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

where $\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. We write $X \sim \mathrm{Beta}(\alpha,\beta)$.

If $X \sim \mathrm{Beta}(\alpha,\beta)$, X has support $(0,1)$, mean $\mu = \frac{\alpha}{\alpha+\beta}$, variance $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ and mgf $M_X(t) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r} \right) \frac{t^k}{k!}$.

2 Multivariate distributions

We will often want to deal with more than one variable based on the same random experiment.

Definition 2.1 (random vector)**.** Consider a random experiment with sample space Ω. Let $X_1, X_2 : \Omega \to \mathbb{R}$ be RVs. We say that (X_1, X_2) is a **random vector**. The support of (X_1, X_2) is the set of ordered pairs $\chi = \{(x_1, x_2) : X_1(\omega) = x_1, X_2(\omega) = x_2, \omega \in \Omega\}$.

Of particular interest are events of the forms $(X_1 \le x_1) \cap (X_2 \le x_2)$ for $(x_1, x_2) \in \mathbb{R}^2$

Definition 2.2 (joint cdf)**.** Suppose that (X_1, X_2) is a random vector. The **joint cdf** is defined as

$$F_{X_1,X_2}(x_1, x_2) = P((X_1 \le x_1) \cap (X_2 \le x_2)).$$

From here we can extend the univariate case for the probability over intervals to rectangular subsets of $\mathbb{R}^2$.

Theorem 2.1. Suppose that the random vector (X_1, X_2) has joint cdf F_{X_1,X_2} and let $a_1, b_1, a_2, b_2 \in \mathbb{R}$ be such that $a_1 < b_1$ and $a_2 < b_2$. Then

$$P((X_1, X_2) \in [a_1, b_1] \times [a_2, b_2]) = F_{X_1,X_2}(b_1, b_2) - F_{X_1,X_2}(a_1, b_2) - F_{X_1,X_2}(b_1, a_2) + F_{X_1,X_2}(a_1, a_2).$$

Recall that, in general, RVs can be of the discrete type or of the continuous type. We extend this idea to random vectors.

A random vector (X_1, X_2) is said to be **discrete** if its support χ is countable. In this case both X_1 and X_2 are discrete RVs. It thus makes sense to define pmfs for random vectors.

Definition 2.3 (joint pmf)**.** Let (X_1, X_2) be a discrete random vector. Then the **joint pmf** of (X_1, X_2) is given by

$$p_{X_1,X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2).$$

As in the univariate case, this joint pmf satisfies the following properties.

1. $0 \le p_{X_1,X_2}(x_1, x_2) \le 1$ for all $(x_1, x_2) \in \chi$.

2. $\sum_{(x_1,x_2)\in\chi} p_{X_1,X_2}(x_1,x_2) = 1.$

Example 2.1. Suppose two dice are rolled. Let X denote the number of dots facing up on the first die and Y the number of dots on the second die. Also, let W denote the larger of the two. We would like to find the joint pmf of (X,W).
By the equally likely model, the joint pmf of (X,W) is summarized by the following table.

W/X	1	2	3	4	5	6
1	$\frac{1}{36}$	0	0	0	0	0
2	$\frac{1}{36}$	$\frac{2}{36}$	0	0	0	0
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{3}{36}$	0	0	0
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{4}{36}$	0	0
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{5}{36}$	0
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$

Note that it is clearly the case that $0 \le p_{X,W}(x,w) \le 1$ and $\sum_{(x,w)\in(X,W)} p_{X,W}(x,w) = 1$.
Suppose we wanted to find

$$\sum_{\substack{(x,w)\in\chi \\ x=w}} p_{X,W}(x,w)$$

$$= \frac{7}{12}$$

If the joint cdf of a random vector (X_1,X_2) is continuous, then we say that (X_1,X_2) is continuous. Similarly to a joint pmf, we can also define a joint pdf

Definition 2.4 (joint pdf)**.** Let (X_1,X_2) be a continuous random vector. Then the **joint pdf** of (X_1,X_2) is the function $f_{X_1,X_2}: \mathbb{R}^2 \to \mathbb{R}_+$ satisfying

$$F_{X_1,X_2}(x_1,x_2) > \int_{-\infty}^{x_2}\int_{-\infty}^{x_1} f_{X_1,X_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v\,.$$

As in the univariate case, the pdf satisfies the following properties

1. $f_{X_1,X_2}(x_1,x_2) \ge 0$ for all $(x_1,x_2)\in\mathbb{R}^2$.
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2}(u,v)\,\mathrm{d}u\,\mathrm{d}v = 1.$

We may also define the marginal cdf of a random vector (X_1,X_2).

$$
\begin{aligned}
F_{X_1} &= P(X_1 \leq x_1) \\
&= P(X_1 \leq x_1, -\infty < X_2 < \infty) \\
&= \lim_{x_2 \to \infty} P(X_1 \leq x_1, X_2 \leq x_2) \\
&= \lim_{x_2 \to \infty} F_{X_1,X_2}(x_1, x_2)
\end{aligned}
$$

We may also find marginal pmfs and pdf.

If (X_1, X_2) is a discrete random vector, then

$$
\begin{aligned}
p_{X_1}(x_1) &= P(X_1 = x_1) \\
&= \sum_{x_2} p_{X_1,X_2}(x_1, x_2)
\end{aligned}
$$

If (X_1, X_2) is a continuous random vector, then

$$
f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, \mathrm{d}x_2
$$

Example 2.2. We revisit the previous example.

W	$p_W(w)$
1	$\frac{1}{36}$
2	$\frac{3}{36}$
3	$\frac{5}{36}$
4	$\frac{7}{36}$
5	$\frac{9}{36}$
6	$\frac{11}{36}$

With such a marginal pmf, we may find marginal expectations directly. In this case, $E(W) = \sum_1^6 w p_W(w)$.

Example 2.3. Consider the joint pdf of (X,Y) to be

$$f_{X,Y}(x,y) = \begin{cases} cxy^2 & \text{for } 0 \le x \le 2, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

We would like to find the value of c such that $f_{X,Y}$ is a valid pdf. We would then like to find the marginal pdfs.
For $f_{X,Y}$ to be valid, we require it to be non-negative and it must integrate to 1. The second property will be used to determine c.

$$\begin{aligned} \int_0^1 \int_0^2 cxy^2 \, dx \, dy &= 1 \\ c\int_0^1 y^2 \int_0^2 x \, dx \, dy &= 1 \\ c\left(\frac{1}{3}\right)(2) &= 1 \\ \frac{2}{3}c &= 1 \\ c &= \frac{3}{2} \end{aligned}$$

Such a c makes $f_{X,Y}$ a valid pdf.
We first find the marginal pdf of X.

$$\begin{aligned} f_X(x) &= \int_0^1 \frac{3}{2}xy^2 \, dy \\ &= \frac{3}{2}x\int_0^1 y^2 \, dy \\ &= \frac{x}{2} \end{aligned}$$

So $f_X(x) = \begin{cases} \frac{x}{2} & \text{for } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$
We leave the marginal pdf of Y as an exercise.

Expectations in the multivariate case are easily extended from the univariate case to random vectors.

Suppose (X_1, X_2) is a random vector and let $Y = g(X_1, X_2)$ for $g : \mathbb{R}^2 \to \mathbb{R}$.

If (X_1, X_2) is discrete with joint pmf p_{X_1,X_2}, then $E(Y)$ exists if [insert condition] and is defined as

$$E(Y) = \sum_{x_1} \sum_{x_2} g(x_1, x_2) p_{X_1,X_2}(x_1, x_2)$$