MATH 170B LECTURE NOTE 0: REVIEW OF MATH 170A

HANBAEK LYU

Many things in life are uncertain. Can we 'measure' and compare such uncertainty so that it helps us to make more informed decision? Probability theory provides a systematic way of doing so.

1. PROBABILITY MEASURE AND PROBABILITY SPACE

We begin with idealizing our situation. Let Ω be a finite set, called *sample space*. This is the collection of all possible outcomes that we can observe (think of six sides of a die). We are going to perform some experiment on Ω, and the outcome could be any subset E of Ω, which we call an *event*. Let us denote the collection of all events $E \subseteq \Omega$ by 2^{Ω}. A *probability measure* on Ω is a function $\mathbb{P}$ such that for each event $E \subseteq \Omega$, it assigns a number $\mathbb{P}(E) \in [0,1]$ and satisfies the following properties:

(i) $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$.
(ii) If two events $E_1, E_2 \subseteq \Omega$ are disjoint, then $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_1)$.

In words, $\mathbb{P}(E)$ is our quantization of how likely it is that the event E occurs out of our experiment.

Exercise 1.1. Let $\mathbb{P}$ be a probability measure on sample space Ω. Show the following.

(i) Let $E = \{x_1, x_2, \cdots x_k\} \subseteq \Omega$ be an event. Then $\mathbb{P}(E) = \sum_{i=1}^{k} \mathbb{P}(\{x_i\}) = 1$.
(ii) $\sum_{x \in \Omega} \mathbb{P}(\{x\}) = 1$.

If $\mathbb{P}$ is a probability measure on sample space Ω, we call the pair $(\Omega, \mathbb{P})$ a *probability space*. This is our idealized world where we can precisely measure uncertainty of all possible events. Of course, there could be many (in fact, infinitely many) different probability measures on the same sample space.

Exercise 1.2 (coin flip)**.** Let $\Omega = \{H, T\}$ be a sample space. Fix a parameter $p \in [0,1]$, and define a function $\mathbb{P}_p : 2^{\Omega} \to [0,1]$ by $\mathbb{P}_p(\emptyset) = 0$, $\mathbb{P}_p(\{H\}) = p$, $\mathbb{P}_p(\{T\}) = 1-p$, $\mathbb{P}_p(\{H,T\}) = 1$. Verify that $\mathbb{P}_p$ is a probability measure on Ω for each value of p.

A typical way of constructing a probability measure is to specify how likely it is to see each individual element in Ω. Namely, let $f : \Omega \to [0,1]$ be a function that sums up to 1, i.e., $\sum_{x \in \Omega} f(x) = 1$. Define a function $\mathbb{P} : 2^{\Omega} \to [0,1]$ by

$$\mathbb{P}(E) = \sum_{\omega \in E} f(\omega). \tag{1}$$

Then this is a probability measure on Ω, and f is called a *probability distribution* on Ω. For instance, the PMF on $\{H, T\}$ we used to define $\mathbb{P}_p$ in Exercise 1.2 is $f(H) = p$ and $f(T) = 1-p$.

Exercise 1.3. Show that the function $\mathbb{P} : 2^{\Omega} \to [0,1]$ defined in (1) is a probability measure on Ω. Conversely, show that every probability measure on a finite sample space Ω can be defined in this way.

Remark 1.4 (General probability space)**.** A probability space does not need to be finite, but we need a more careful definition in that case. For example, if we take Ω to be the unit interval $[0,1]$, then we have to be careful in deciding which subset $E \subseteq \Omega$ can be an 'event': not every subset of Ω can be an event. A proper definition of general probability space is out of the scope of this course.

2. (DISCRETE) RANDOM VARIABLES

Given a probability space $(\Omega, \mathbb{P})$, a *random variable* (RV) is a real-valued function $X : \Omega \to \mathbb{R}$. We can think of it as the outcome of some experiment on Ω (e.g., height of a randomly selected friend). We often forget the original probability space and specify a RV by is *probability mass function* (PMF) $f_X : \mathbb{R} \to [0,1]$,

$$f_X(x) = \mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega \,|\, X(\omega) = x\}). \tag{2}$$

Namely, $\mathbb{P}(X = x)$ is the likelihood that the RV X takes value x.

Example 2.1. Say you win \$1 if a fair coin lands heads and lose \$1 if lands tails. We can set up our probability space $(\Omega, \mathbb{P})$ by $\Omega = \{H, T\}$ and $\mathbb{P} = \mathbb{P}_{1/2}$ as in Exercise 1.2. The RV $X : \Omega \to \mathbb{R}$ for this game is $X(H) = 1$ and $X(T) = -1$. The PMF of X is given by $f_X(1) = \mathbb{P}(X = 1) = \mathbb{P}(\{H\}) = 1/2$ and likewise $f_X(-1) = 1/2$.

Exercise 2.2. Let $(\Omega, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ be a RV. Show that its PMF f_X adds up to 1, that is,

$$\sum_x f_X(x) = 1, \tag{3}$$

where the summation runs over all numerical values x that X can take.

There are two useful statistics of a RV to summarize its property. First, if one has to guess the value of a RV X, what would be the best choice? It is the *expectation* (or mean) of X, defined as below:

$$\mathbb{E}(X) = \sum_x x\,\mathbb{P}(X = x). \tag{4}$$

Exercise 2.3 (Tail sum formula for expectation)**.** For any RV X taking values on positive integers, show that

$$\mathbb{E}(X) = \sum_{x=1}^{\infty} \mathbb{P}(X \geq x). \tag{5}$$

Exercise 2.4 (Linearity of expectation)**.** In this exercise, we will show that the expectation of sum of RVs is the sum of expectation of individual RVs.

(i) Let X and Y be RVs. Show that

$$\sum_y \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x). \tag{6}$$

(ii) Verify the following steps:

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_z z\,\mathbb{P}(X+Y=z) && (7)\\
&= \sum_z \sum_{\substack{x,y\\ x+y=z}} (x+y)\mathbb{P}(X=x,\, Y=y) && (8)\\
&= \sum_{x,y} (x+y)\,\mathbb{P}(X=x,\, Y=y) && (9)\\
&= \sum_{x,y} x\,\mathbb{P}(X=x,\, Y=y) + \sum_{x,y} y\,\mathbb{P}(X=x,\, Y=y) && (10)\\
&= \sum_x x\left(\sum_y \mathbb{P}(X=x, y=y)\right) + \sum_y y\left(\sum_x \mathbb{P}(X=x,\, Y=y)\right) && (11)\\
&= \sum_x x\mathbb{P}(X=x) + \sum_y y\mathbb{P}(Y=y) && (12)\\
&= \mathbb{E}(X) + \mathbb{E}(Y). && (13)
\end{aligned}$$

(iii) Use induction to show that for any RVs $X_1, X_2, \cdots, X_n$, we have

$$\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n). \tag{14}$$

On the other hand, say you play two different games where in the first game, you win or lose \$1 depending on a fair coin flip, and in the second game, you win or lose \$10. In both games, your expected winning is 0. But the two games are different in how much the outcome fluctuates around the mean. This notion if fluctuation is captured by the following quantity called *variance*:

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]. \tag{15}$$

Namely, it is the expected squared difference between X and its expectation $\mathbb{E}(X)$.

Exercise 2.5. For any RV X, show that

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \tag{16}$$

Exercise 2.6. In this exercise, we will see how we can express the variance of sums of RVs. For two RVs X and Y, define their *covariance* $\mathrm{Cov}(X, Y)$ by

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \tag{17}$$

(i) Use Exercises 2.5 and 2.4 to show that

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y). \tag{18}$$

(ii) Use induction to show that for RVs $X_1, X_2 \cdots, X_n$

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \le i, j \le n} \mathrm{Cov}(X_i, X_j) \tag{19}$$

Here are some of the simplest and yet most important RVs.

Exercise 2.7. (Bernoulli RV) A RV X is a *Bernoulli* variable with (success) probability $p \in [0, 1]$ if it takes value 1 with probability p and 0 with probability $1 - p$. In this case we write $X \sim \mathrm{Bernoulli}(p)$. Show that $\mathbb{E}(X) = p$ and $\mathrm{Var}(X) = p(1 - p)$.

Exercise 2.8 (Indicator variables)**.** Let $(\Omega, \mathbb{P})$ be a probability space and let $E \subseteq \Omega$ be an event. The *indicator variable* of the event E, which is denoted by $\mathbf{1}_E$, is the RV such that $\mathbf{1}_E(\omega) = 1$ if $\omega \in E$ and $\mathbf{1}_E(\omega) = 0$ if $\omega \in E^c$. Show that $\mathbf{1}_E$ is a Bernoulli variable with success probability $p = \mathbb{P}(E)$.

3. Conditioning and independence

Consider two experiments on a probability space and the outcomes are recorded by RVs X and Y. For instance, X could be the number of friends on Facebook and Y could be the number of connections on LinkedIn of a randomly chosen classmate. Perhaps it would be case that Y is large if X is large. Or maybe the opposite is true. In any case, the outcome of Y is most likely be affected by knowing something about X. This leads to the notion of 'conditioning'. For any two events E_1 and E_2 such that $\mathbb{P}(E_2) > 0$, we define

$$\mathbb{P}(E_1 \mid E_2) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_2)} \tag{20}$$

and this quantity is called the *conditional probability* of E_1 given E_2. For RVs X, Y and subsets $A_1, A_2 \subseteq \mathbb{R}$, we similarly define

$$\mathbb{P}(X \in A_1 \mid Y \in A_2) = \frac{\mathbb{P}(X \in A_1 \text{ and } Y \in A_2)}{\mathbb{P}(Y \in A_2)}. \tag{21}$$

This is the conditional probability that X belongs to A_1 given that Y belongs to A_2. The *conditional expectation* of X given $Y = y$ is defined by

$$\mathbb{E}(X \mid Y = y) = \sum_{x} x \mathbb{P}(X = x \mid Y = y). \tag{22}$$

Example 3.1. Consider two RVs X and Y taking values from $\{0,1,2,3\}$. Their *joint PMF* is depicted in Figure 1. Then

$$\begin{aligned}\mathbb{P}(X \geq 2 \,|\, Y=2) &= \mathbb{P}(X=2 \,|\, Y=2) + \mathbb{P}(X \geq 3 \,|\, Y=2) && (23)\\ &= \frac{\mathbb{P}(X=2 \text{ and } Y=2)}{\mathbb{P}(Y=2)} + \frac{\mathbb{P}(X=3 \text{ and } Y=2)}{\mathbb{P}(Y=2)} && (24)\\ &= \frac{5/33}{(3+0+5+1)/33} + \frac{1/33}{(3+0+5+1)/33} = 2/3. && (25)\end{aligned}$$

Moreover,

$$\begin{aligned}\mathbb{E}(X \,|\, Y=2) &= \sum_{x=0}^{3} x\,\mathbb{P}(X=x \,|\, Y=2) && (26)\\ &= 0\frac{1/33}{9/33} + 1\frac{0}{9/33} + 2\frac{5/33}{9/33} + 3\frac{1/33}{9/33} = 13/9. && (27)\end{aligned}$$

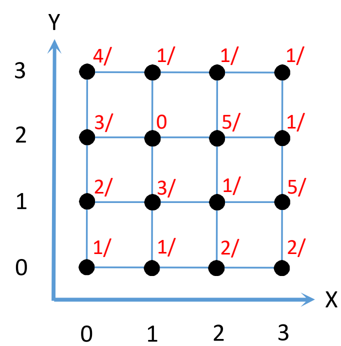

FIGURE 1. Two RVs X, Y and their joint distribution in red. Common denominator of 33 is omitted in the figure.

When knowing something about one RV does not yield any information of the other, we say the two RVs are independent. Formally, we say two events E_1 and E_2 are *independent* if

$$\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2). \tag{28}$$

Two RVs X and Y are *independent* if for any two subsets $A_1, A_2 \subseteq \mathbb{R}$,

$$\mathbb{P}(X \in A_1 \text{ and } Y \in A_2) = \mathbb{P}(X \in A_1)\mathbb{P}(Y \in A_2). \tag{29}$$

We say two events or RVs are *dependent* if they are not independent.

Exercise 3.2. Suppose two RVs X and Y are independent. Then for any subsets $A_1, A_2 \subseteq \mathbb{R}$ such that $\mathbb{P}(Y \in A_2) > 0$, show that

$$\mathbb{P}(X \in A_1 \,|\, Y \in A_2) = \mathbb{P}(X \in A_1). \tag{30}$$

Example 3.3. Flip two fair coins at the same time, and let $X = 1$ if the first coin lands heads and $X = -1$ if it lands tails. Let Y be a similar RV for the second coin. Clearly knowing about one coin does not give any information of the other. For instance, the first coin lands on heads with probability $1/2$. Whether the first coin lands on heads or not, the second coin will land on heads with probability $1/2$. So

$$\mathbb{P}(X=1 \text{ and } Y=1) = \frac{1}{2}\cdot\frac{1}{2} = \mathbb{P}(X=1)\mathbb{P}(Y=1). \tag{31}$$

Exercise 3.4. Recall the definition of covariance given in Exercise 2.6.

(i) Show that if two RVs X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$

(ii) Use Exercise 2.6 to conclude that if $X_1, \cdots, X_n$ are independent RVs, then

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n). \tag{32}$$

4. Binomial, geometric, and Poisson RVs

Example 4.1 (Binomial RV)**.** Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed Bernoulli p variables. Let $X = X_1 + \cdots + X_n$. One can think of flipping the same probability p coin n times. Then X is the total number of heads. Note that X has the following PMF

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^k \tag{33}$$

for k nonnegative integer, and $\mathbb{P}(X = k) = 0$ otherwise. We say X follows the Binomial distribution with parameters n and p, and write $X \sim \mathrm{Binomial}(n, p)$.

We can compute the mean and variance of X using the above PMF directly, but it is much easier to break it up into Bernoulli variables and use linearity. Recall that $X_i \sim \mathrm{Bernoulli}(p)$ and we have $\mathbb{E}(X_i) = p$ and $\mathrm{Var}(X_i) = p(1-p)$ for each $1 \le i \le n$ (from Exercise 2.7). So by linearity of expectation (Exercise 2.4),

$$\mathbb{E}(X) = \mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np. \tag{34}$$

On the other hand, since X_i's are independent, variance of X is the sum of variance of X_i's (Exercise 3.4) so

$$\mathrm{Var}(X) = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = np(1-p). \tag{35}$$

Example 4.2 (Geometric RV)**.** Suppose we flip a probability p coin until it lands heads. Let X be the total number of trials until the first time we see heads. Then in order for $X = k$, the first $k-1$ flips must land on tails and the kth flip should land on heads. Since the flips are independent with each other,

$$\mathbb{P}(X = k) = \mathbb{P}(\{T, T, \cdots, T, H\}) = (1-p)^{k-1} p. \tag{36}$$

This is valid for k positive integer, and $\mathbb{P}(X = k) = 0$ otherwise. Such a RV is called a *Geometric RV* with (success) parameter p, and we write $X \sim \mathrm{Geom}(p)$.

The mean and variance of X can be easily computed using its moment generating function, which we will learn soon in this course. For their direct computation, note that

$$\begin{aligned}
\mathbb{E}(X) - (1-p)\mathbb{E}(X) &= (1-p)^0 p + 2(1-p)^1 p + 3(1-p)^2 p + 4(1-p)^3 p \cdots && (37)\\
&\quad - \left[(1-p)^1 p + 2(1-p)^2 p + 3(1-p)^3 p + \cdots\right] && (38)\\
&= (1-p)^0 p + (1-p)^1 p + (1-p)^2 p + (1-p)^3 p \cdots && (39)\\
&= \frac{p}{1-(1-p)} = 1, && (40)
\end{aligned}$$

where we recognized the series after the second equality as a geometric series. This gives

$$\mathbb{E}(X) = 1/p. \tag{41}$$

Exercise 4.3. Let $X \sim \mathrm{Geom}(p)$. Use a similar computation as we had in Example 4.2 to show $\mathbb{E}(X^2) = (2-p)/p^2$. Using the fact that $\mathbb{E}(X) = 1/p$, conclude that $\mathrm{Var}(X) = (1-p)/p^2$.

Example 4.4 (Poisson RV)**.** A RV X is a *Poisson RV* with rate $\lambda > 0$ if

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{42}$$

for all nonnegative integers $k \ge 0$. We write $X \sim \mathrm{Poisson}(\lambda)$.

Poisson distribution is obtained as a limit of the Binomial distribution as the number n of trials tend to infinity while the mean np is kept at constant λ. Namely, let $Y \sim \text{Binomial}(n,p)$ and suppose $np = \lambda$. This means that we expect to see λ successes out of n trials. Then what is the probability that we see, say, k successes out of n trials, when n is large? Since the mean is λ, this probability should be very small when k is large compared to λ. Indeed, we can rewrite the Binomial PMF as

$$\mathbb{P}(Y=k) = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} \tag{43}$$

$$= \frac{n}{n}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\frac{(np)^k}{k!}(1-p)^{n-k} \tag{44}$$

$$= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k}. \tag{45}$$

As n tends to infinity, the limit of the last expression is precisely the right hand side of (42). [1]

Exercise 4.5. Let $X \sim \text{Poisson}(\lambda)$. Show that $\mathbb{E}(X) = \text{Var}(X) = \lambda$.

5. Continuous RVs

So far we have only considered discrete RVs, which takes either finitely many or countably many values. While there are many examples of discrete RVs, there are also many instances of RVs which various continuously (e.g., temperature, height, weight, price, etc.). To define a discrete RV, it was enough to specify its PMF. For a continuous RV, *probability distribution function* (PDF) plays an analogous role of PMF. We also need to replace summation $\sum$ with an integral $\int dx$.

Namely, X is a continuous RV if there is a function $f_X : \mathbb{R} \to [0,\infty)$ such that for any interval $[a,b]$, the probability that X takes a value from an interval $(a,b]$ is given by integrating f_X over the interval $(a,b]$:

$$\mathbb{P}(X \in (a,b]) = \int_a^b f_X(x)\,dx. \tag{46}$$

The *cumulative distribution function* (CDF) of a RV X (either discrete or continuous), denoted by F_X, is defined by

$$F_X(x) = \mathbb{P}(X \le x). \tag{47}$$

By definition of PDF, we get

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt. \tag{48}$$

Conversely, PDFs can be obtained by differentiating corresponding CDFs.

Exercise 5.1. Let X be a continuous RV with PDF f_X. Let a be a continuity point of f_X, that is, f_X is continuous at a. Show that $F_X(x)$ is differentiable at $x = a$ and

$$\left.\frac{dF_X}{dx}\right|_{x=a} = f_X(a). \tag{49}$$

The expectation of a continuous RV X with pdf f_X is defined by

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx, \tag{50}$$

and its variance $\text{Var}(X)$ is defined by the same formula (15).

[1] Later, we will interpret the value of a Poisson variable $X \sim \text{Poisson}(\lambda)$ as the number of customers arriving during a unit time interval, where the waiting time between consecutive customers is distributed as an independent exponential distribution with mean $1/\lambda$. Such an arrival process is called the Poisson process.

Exercise 5.2. (Tail sum formula for expectation) Let X be a continuous RV with PDF f_X and suppose $f_X(x) = 0$ for all $x < 0$. Use Fubini's theorem to show that

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq t)\, dt. \tag{51}$$

6. Uniform, exponential, and normal RVs

Example 6.1 (Uniform RV)**.** X is a *uniform RV* on the interval $[a,b]$ (denoted by $X \sim \text{Uniform}([a,b])$) if it has PDF

$$f_X(x) = \frac{1}{b-a}\mathbf{1}(a \leq x \leq b). \tag{52}$$

An easy computation gives its CDF:

$$\mathbb{P}(X \leq x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \leq x \leq b \\ 1 & x > b. \end{cases} \tag{53}$$

Exercise 6.2. Let $X \sim \text{Uniform}([a,b])$. Show that

$$\mathbb{E}(X) = \frac{a+b}{2}, \qquad \text{Var}(E) = \frac{(b-a)^2}{12}. \tag{54}$$

Example 6.3 (Exponential RV)**.** X is an *exponential* RV with rate λ (denoted by $X \sim \text{Exp}(\lambda)$) if it has PDF

$$f_X(x) = \lambda e^{-\lambda x}\mathbf{1}(x \geq 0). \tag{55}$$

Integrating the PDF gives its CDF

$$\mathbb{P}(X \leq x) = (1 - e^{-\lambda x})\mathbf{1}(x \geq 0). \tag{56}$$

Using Exercise 5.2, we can compute

$$\mathbb{E}(X) = \int_0^\infty e^{-\lambda t}\, dt = \left[-\frac{e^{-\lambda t}}{\lambda}\right]_0^\infty = 1/\lambda. \tag{57}$$

Exercise 6.4. Let $X \sim \text{Exp}(\lambda)$. Show that $\mathbb{E}(X) = 1/\lambda$ directly using definition (50). Also show that $\text{Var}(X) = 1/\lambda^2$.

Example 6.5. 1.28[Normal RV] X is a *normal* RV with mean μ and variance σ^2 (denoted by $X \sim N(\mu, \sigma^2)$) if it has PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{58}$$

If $\mu = 0$ and $\sigma^2 = 1$, then X is called a standard normal RV. Note that if $X \sim N(\mu, \sigma^2)$, then $Y := X - \mu$ has PDF

$$f_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}. \tag{59}$$

Since this is an even function, it follows that $\mathbb{E}(Y) = 0$. Hence $\mathbb{E}(X) = \mu$.

Exercise 6.6 (Gaussian integral)**.** In this exercise, we will show $\int_{-\infty}^\infty e^{-x^2}\, dx = \sqrt{\pi}$.

(i) Show that

$$\int x e^{-x^2}\, dx = -\frac{1}{2} e^{-x^2} + C. \tag{60}$$

(ii) Let $I = \int_{-\infty}^\infty e^{-x^2}\, dx$. Show that

$$I^2 = \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-(x^2+y^2)}\, dx\, dy. \tag{61}$$

(iii) Use polar coordinate (r,θ) to rewrite

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2}\, r\, dr\, d\theta = 2\pi \int_0^{\infty} r e^{-r^2}\, dr. \tag{62}$$

Then use (i) to deduce $I^2 = \pi$. Conclude $I = \sqrt{\pi}$.

Exercise 6.7. Let $X \sim N(\mu, \sigma^2)$. In this exercise, we will show $\mathrm{Var}(X) = \sigma^2$.

(i) Show that $\mathrm{Var}(X) = \mathrm{Var}(X - \mu)$.

(ii) Use integration by parts and Exercise 6.6 to show that

$$\int_0^{\infty} x^2 e^{-x^2}\, dx = \left[x \left(-\frac{1}{2} e^{-x^2} \right) \right]_0^{\infty} + \int_0^{\infty} \frac{1}{2} e^{-x^2}\, dx = \frac{\sqrt{\pi}}{4}. \tag{63}$$

(iii) Use change of variable $t = \sqrt{2}\sigma x$ and (ii) to show

$$\int_0^{\infty} \frac{x^2}{\sqrt{2\pi\sigma^2}} e^{-x^2}\, dx = \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{\infty} t^2 e^{-t^2}\, dt = \frac{\sigma^2}{2}. \tag{64}$$

Conclude $\mathrm{Var}(X) = \sigma^2$.