MATH 157: Intermediate Probability

Connor Neely, Fall 2024

* Adapted from FA24 Lecture Notes.

1 Univariate Distributions

1.1 Discrete Random Variables

The fundamental object of this class is the random variable. We'll start with the discrete case.

Definition: Probability mass function

Let X be a discrete random variable. The function $P(X = k)$ is called a probability mass function if

- $P(X = k) \geq 0$ for all k and
- $\sum_k P(X = k) = 1$.

Before looking at some important kinds of probability mass functions, we'll examine some of their general characteristics regarding center and spread.

Definition: Expected value

The expected value of a discrete random variable X is given by

$$E(X) = \sum_k k\, P(X = k).$$

The expected value of X is also called the mean of X, and is sometimes denoted μ or μ_X. Also, for any function $g(x)$ we define

$$E[g(x)] = \sum_k g(k)\, P(X = k).$$

Note that the expected value may not exist if the above sums do not converge.

Definition: Variance and standard deviation

The variance of a random variable X with mean μ is the average squared distance from μ. That is,

$$\mathrm{var}(X) = E[(x - \mu)^2].$$

It is typically denoted by σ^2 or σ_x^2. The standard deviation is

$$\mathrm{sd}(x) = \sqrt{\mathrm{var}(x)},$$

and is typically denoted by σ or σ_x.

There's a couple of theorems regarding expected value and variance which will prove useful in computations.

Theorem 1.1: Expected value and variance of linear expressions

For real a and b,

$$E(ax + b) = a\, E(x) + b \quad \text{and} \quad \mathrm{var}(ax + b) = a^2\, \mathrm{var}(x).$$

Theorem 1.2: Variance via expected value

For a random variable X,

$$\mathrm{var}(X) = E(X^2) - E(X)^2.$$

We'll prove later that a random variable is almost always within two standard deviations of its mean, with probability greater at least 75%.

1.2 Some Important Discrete RVs

The simplest kind of random variable is a uniform random variable, in which all outcomes have equal probabilities. This is boring, though, so let's define a few more interesting ones.

Definition: Geometric random variable

Consider an infinite sequence of independent binary events, each of which has a probability p of success. If X is the index of the first success, then X is called a geometric random variable and

$$P(X = k) = (1 - p)^{k-1}p.$$

We say that X obeys a geometric distribution with parameter p, or $X \sim \text{geo}(p)$.

Theorem 1.3: EV of a geometric random variable

A random variable $X \sim \text{geo}(p)$ has expected value $E(X) = 1/p$.

We omit a proof because the result is intuitive, but it involves differentiating the geometric series sum.

As a side note, we could compute $P(X \geq k)$ in a few ways: via a finite geometric series, by subtracting off $P(X < k)$, or by finding the probability $(1-p)^{k-1}$ that the first $k-1$ events are failures. Now onto another!

Definition: Binomial random variable

Consider a sequence of n independent binary events, each of which has a probability p of success. If X is the number of successes in this sequence, then X is called a binomial random variable and

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

We say that X obeys a geometric distribution with parameters n and p, or $X \sim \text{bin}(n, p)$.

Theorem 1.4: EV and variance of a binomial random variable

A random variable $X \sim \text{bin}(n, p)$ has $E(X) = np$ and $\text{var}(X) = np(1 - p)$.

Another intuitive result, and we'll defer a proof to later. For our last variable, we start with $X \sim \text{bin}(n, p)$ and take the large and small limits of n and p, respectively. This gives us the following.

Definition: Poisson random variable

Consider a long sequence of rare binary events with expected value $\lambda > 0$. If X is the number of successes in the sequence, then X is called a Poisson random variable and

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

We say that X obeys a Poisson distribution with parameter λ, or $X \sim \text{poi}(\lambda)$.

Theorem 1.5: Variance of a Poisson random variable

A random variable $X \sim \text{poi}(\lambda)$ has $\text{var}(X) = \lambda$.

The full proof of this is lengthy. As a jumping-off point, it involves computing $E(X^2) = E[X(X - 1) + X]$.

1.3 Conditional Probability

We'll finish off our discrete discussion by looking at conditional probability.

Definition: Conditional probability

For two events A and B, the probability of A given B is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

From this definition we can write a few intuitive identities.

Theorem 1.6: Conditional probability identities

For two events A and B, the following are true.

(a) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(b) $P(A \cap B) = P(A)P(B)$ for independent A, B.

(c) $P(A) = P(A \cap B) + P(A \cap B^c)$.

Theorem 1.7: Law of total probability

For any A and disjoint $B_1, \cdots, B_n$,

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i).$$

Everything here is pretty straightforward, but now we'll look at how it all ties into more complicated scenarios.

Example: Craps

Suppose we roll two dice to get a total X. There are three outcomes: if $X = 7, 11$ then we win immediately; if $X = 2, 3, 12$ we lose immediately; and if X is anything else we continue rolling until X appears again, in which case we win, or until 7 appears, in which case we lose. The probability of winning is

$$P(\text{win}) = \sum_{k=2}^{12} P(\text{win} \mid X = k)\, P(X = k).$$

This is easy enough for the first two cases, so we'll focus on the third. In particular, we can compute $P(\text{win} \mid X = 4)$ in three different ways. Let Π_4 denote the probability of rolling another 4 before a 7.

- We could simply use a geometric series:

$$\Pi_4 = \frac{3}{36} + \left(\frac{27}{36}\right)\left(\frac{3}{36}\right) + \left(\frac{27}{36}\right)^2 \left(\frac{3}{36}\right) + \cdots = \frac{1}{3}.$$

- We could also solve the equation

$$\Pi_4 = \frac{3}{36} + \frac{27}{36}\Pi_4 \implies \Pi_4 = \frac{1}{3}.$$

- Finally, we could simply observe that there are three ways to roll a 4 and six ways to roll a 7. Thus $\Pi_4 = 3/(3+6) = 1/3$.

Doing the same for all of the other possibilities will yield $P(\text{win}) = 0.49\overline{29}$.

Craps is very close to being a fair game, but for repeated bets the odds are deceptive.

Theorem 1.8: Gambler's ruin

Consider a game with probability p of success. A player making repeated \$1 bets starts with \$$i$ and aims to reach \$$n$, with $0 \leq i \leq n$. The probability of reaching this goal is given by

$$a_i = \frac{1-(q/p)^i}{1-(q/p)^n},$$

where $p \neq 1/2$ and $q = 1-p$.

Proving this involves solving the recurrence

$$a_i = pa_{i+1} + qa_{i-1}, \quad a_0 = 0, \ a_n = 1;$$

For a fair game ($p = 1/2$) we get $a_i = i/n$, and in any other case with $p \neq 0$ we get the result above.

1.4 Continuous Random Variables

Now we'll make our discussion a bit more interesting by considering the continuous case.

Definition: Probability density function

A continuous random variable is described by a probability density function f_X (or just f or $f(x)$) such that, for any $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx.$$

Such a function must satisfy $f_X(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

The key characteristics we have for discrete random variables generalize nicely to continuous ones. Note that the expected value may not exist if the distribution has "heavy tails", in which case the variance also doesn't exist.

Definition: Expected value

For a continuous random variable X,

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

The variance and standard deviation are the same as we defined before: $\mathrm{var}(X) = E[(x-\mu)^2]$, where μ is the distribution's mean. Also, all of the nice properties we found about these characteristics hold, too! Namely:

$$E(ax+b) = a\,E(x) + b, \qquad \mathrm{var}(ax+b) = a^2 \mathrm{var}(x), \qquad \mathrm{var}(X) = E(X^2) - E(X)^2.$$

By the linearity of integration we could generalize the first result to arbitrary linear combinations of functions. Now we'll make one more definition that we really could've made earlier in the discrete case but also wouldn't have been useful until now.

Definition: Cumulative distribution function

For any random variable X, we define the cumulative distribution function

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

Note that for discrete X this is defined as a summation, while for continuous X it's an integral.

These functions are useful not only in their own right, but also in working with functions of a random variable: given a pdf for X and an expression for Y in terms of X, we can easily generate a pdf for Y. This is best illustrated via an example.

Example: A function of a random variable

Consider a continuous random variable with pdf $f_X(x)$ defined on $0 \le x \le 30$, and let $Y = \frac{9}{5}X + 32$. To write down a pdf for Y we first compute its feasible region:

$$0 \le x \le 30 \implies 32 \le \frac{9}{5}x + 32 \le 86 \implies 32 \le y \le 86.$$

Now we look to the cumulative distribution functions, writing $F_Y(y)$ in terms of F_X:

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P\left(\frac{9}{5}X + 32 \le y\right) \\ &= P\left(X \le \frac{5}{9}(y-32)\right) \\ &= F_X\left(\frac{9}{5}(y-32)\right). \end{aligned}$$

Finally, we can differentiate both sides with respect to y to get the pdf

$$f_Y(y) = \frac{9}{5} f_X\left(\frac{9}{5}(y-32)\right).$$

1.5 Some Important Continuous RVs

Now we'll look at some particular examples of continuous random variables. Like before, we can start simple with the uniform random variable.

Definition: Uniform random variable

X is a uniform random variable with parameters $a < b$ if it has a pdf

$$f(x) = \frac{1}{b-a}.$$

Such a random variable is denoted $X \sim u(a, b)$.

Theorem 1.9: EV and variance of a uniform RV

If $X \sim u(a, b)$ then

$$E(X) = \frac{a+b}{2}, \qquad \operatorname{var}(X) = \frac{(b-a)^2}{12}.$$

What's interesting about this is that we might treat an arbitrary distribution function $F_X(x)$ as a probability distribution, note that it ranges from 0 to 1, and find that $F_X(x) \sim u(0, 1)$. We might restate this in a slightly different way to make it more intuitive.

Theorem 1.10

If $U \sim u(0, 1)$, then $y = F_X^{-1}(U)$ has the same pdf as X. (It follows that a number U generated from $u(0, 1)$ represents the $100u$-th percentile of X.)

Next, we have the continuous analog of the geometric distribution.

Definition: Exponential random variable

X is an exponential random variable with parameter $\lambda > 0$ if X has pdf

$$f(x) = \lambda e^{-\lambda x}.$$

Such a random variable is denoted $X \sim \mathrm{expo}(\lambda)$.

λ is often called the rate parameter, and $1/\lambda$ represents the "average time" between rare events. In this way, the λ here is the same as in the Poisson distribution!

Theorem 1.11: EV and variance of an exponential RV

If $X \sim \mathrm{expo}(\lambda)$, then

$$E(X) = \frac{1}{\lambda}.$$

Also, $E(X^2) = 2/\lambda^2$, so

$$\mathrm{var}(X) = E(X^2) - E(X)^2 = \frac{1}{\lambda^2}.$$

Finally, we have what is perhaps the most important random variable of all. Given its significance, we'll take a slightly closer look at its key properties.

Definition: Normal random variable

X is a normal random variable with parameters μ and σ^2 if it has pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty.$$

Such a random variable is denoted $X \sim n(\mu, \sigma^2)$.

This is a complicated pdf! To simplify it slightly, we could use the change of variables $z = (x-\mu)/\sigma$ to get

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

This is the pdf of the standard normal, $Z \sim n(0,1)$; we call it $\phi(z)$. It still has no elementary antiderivative, though, so in practice we map problems to this standard normal's cdf and use numerical methods to calculate any probabilities. (It's useful to bear in mind that approximately 68% of the distribution is contained within σ of the mean, approximately 95% within 2σ, and 99.7% within 3σ.)

Theorem 1.12: Transforming normal RVs

If $X \sim n(\mu, \sigma^2)$ and $Y = aX + b$, then

(a) $E(Y) = a\mu + b$ and

(b) $\mathrm{var}(Y) = a^2\sigma^2$.

In fact, Y is also normal with these parameters.

Later we'll also prove that linear combinations of independent normal RVs X_i are, in a way, "preserved"; in particular,

$$\sum_i (a_i X_i + b) \sim n\left(\sum_i (a_i \mu_i + b),\ \sum_i a_i^2 \sigma_i^2\right).$$

Importantly, if we have two identical but independent RVs X_1, X_2, something like $X_1 + X_2$ does not have the same distribution as $2X_1$—their variances are different!